

BEST AVAILABLE COPY

PCT
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/02, 15/53, 15/29	A1	(11) International Publication Number: WO 9820113
(21) International Application Number: PCT/US97/0391	(43) International Publication Date: 14 May 1998 (14.05.98)	(44) International Filing Date: 7 November 1997 (07.11.97)
(22) International Priority Date: 8 November 1996 (08.11.96) US 60/035,222 60/044,380	31 July 1997 (31.07.97) US	(31) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SR, SG, SI, SK, SL, TI, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TI, TZ), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG),
(71) Applicant (for all designated States except US): WASHINGTON STATE UNIVERSITY RESEARCH FOUNDATION (USA); N.E. 1615 E ligate Boulevard, Pullman, WA 99164-1802 (US).		
(72) Inventors, and (73) Inventors/Applicants (for US only): LEWIS, Norman, G. (USA); DAVIN, Laurence, B. (IFR/US); DINKOVA-KOSTOVA, Alberta, T. (BG/US); Broadview Apartments #1025, 116 W. University Parkway, Bellevue, MD 21210 (US); FUJITA, Matsuyuki (JP/US); 215 N.W. Anthony #3, Pullman, WA 99163 (US); GANG, David, R. (US/US); 911 S.E. Kleengard Avenue, B-4, Pullman, WA 99163 (US); SARKANEN, Sino (FI/US); 4051 - 39th Avenue S., Minneapolis, MN 55406 (US).		
		Published With International search report.

(54) Title: RECOMBINANT PINOSYNOULARICRESINOL REDUCTASE, RECOMBINANT DIRIGENT PROTEIN, AND METHODS OF USE

(57) Abstract

Dirigent proteins and pinosyntol/laricresinol reductases have been isolated from *Forstertia intermedia*, *Thiolla plicata* and *Tulipa heterophylla*, together with cDNAs encoding dirigent proteins and pinosyntol/laricresinol reductases from these species. Accordingly, isolated DNA sequences are provided which code for the expression of dirigent proteins and pinosyntol/laricresinol reductases. In other aspects, replicative recombinant cloning vehicles are provided which code for dirigent proteins or pinosyntol/laricresinol reductases or for a base sequence sufficiently complementary to at least a portion of a dirigent protein or pinosyntol/laricresinol reductase DNA or RNA to enable hybridization therewith (e.g., antisense dirigent protein or pinosyntol/laricresinol reductase RNA or fragments of complementary dirigent protein or pinosyntol/laricresinol reductase DNA which are useful as polymerase chain reaction primers or as probes for genes encoding dirigent proteins or pinosyntol/laricresinol reductases). In yet other aspects, modified host cells are provided that have been transformed, transfected, infected and/or injected with a recombinant cloning vehicle and/or DNA sequence encoding dirigent protein or pinosyntol/laricresinol reductases. Thus, systems and methods are provided for the recombinant expression of dirigent proteins and/or pinosyntol/laricresinol reductases that may be used to facilitate the production, isolation and purification of significant quantities of recombinant dirigent proteins and/or pinosyntol/laricresinol reductases for subsequent use, to obtain expression or enhanced expression of dirigent proteins and/or pinosyntol/laricresinol reductases in plants in order to enhance, or otherwise alter, lignan biosynthesis, or may be otherwise employed for the regulation or expression of dirigent proteins and pinosyntol/laricresinol reductases.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.	
AL	Spain
AM	ES
AT	FI
AU	Finland
AU	FR
AU	France
AU	GA
AU	Gabon
AU	GB
AU	United Kingdom
AU	GH
AU	Ghana
AU	GN
AU	Guinea
AU	GR
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Korea, People's
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CU
AU	Cuba
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland
AU	GB
AU	Greece
AU	HK
AU	Hong Kong
AU	IL
AU	Israel
AU	IS
AU	Iceland
AU	IT
AU	Italy
AU	JP
AU	Japan
AU	KR
AU	Kazakhstan
AU	KU
AU	China
AU	CZ
AU	Czech Republic
AU	DE
AU	Germany
AU	DK
AU	Denmark
AU	ES
AU	Ecuador
AU	FR
AU	Finland</td

RECOMBINANT PINORESINOL/LARICRESINOL REDUCTASE, RECOMBINANT DIRIGENT PROTEIN AND METHODS OF USE

卷之三

The present invention relates to isolated dirigent proteins and pinoresinol/laricresinol reductases from *Forsythia intermedia*, *Tsuga heterophylla* and *Thuya plicata*, to nucleic acid sequences which code for dirigent proteins and pinoresinol/laricresinol reductases from *Forsythia intermedia*, *Tsuga heterophylla* and *Thuya plicata*, and to vectors containing the sequences, host cells containing the sequences and methods of producing recombinant pinoresinol/laricresinol reductases, recombinant dirigent protein and their mutants.

Background of the Invention

Lignans are a large, structurally diverse, class of vascular plant metabolites having a wide range of physiological functions and pharmacologically important properties (Ayers, D.C., and Loike, J.D. in *Chemistry and Pharmacology of Natural Products. Lignans. Chemical, Biological and Clinical Properties*, Cambridge University Press, Cambridge, England (1990); Lewis et al., in Chemistry of the Amazon, Biodiversity Natural Products, and Environmental Issues, 588, (P.R. Seidl, O.R. Gottlieb and M.A.C. Kaplan) 135-167, ACS Symposium Series, Washington D.C. (1995)). Because of their pronounced antibiotic properties (Markkanen, T. et al., *Drugs Expl. Clin. Res.* 7:71-718 (1981)), antioxidant properties (Fauré, M. et al., *Phytochemistry* 29:3773-3775 (1990); Osawa, T. et al., *Agric. Biol. Chem.* 49:3351-3352 (1985)) and antifeedant properties (Harmatha, J., and Navrot, J., *Biochem. Syst. Ecol.* 12:95-98 (1984)), a major role of lignans in vascular plants is to help confer resistance against various opportunistic biological

pathogens and predators. Lignans have also been proposed as cytokinins (Binns, A.N. et al., *Proc. Natl. Acad. Sci. USA* 84:980-984 (1987)) and as intermediates in lignification (Fahman, M.M.A. et al., *Phytochemistry* 29:1861-1866 (1990)), suggesting a critical role in plant growth and development. It is widely held that elaboration of biochemical pathways to lignins/lignans and related substances from phenylalanine (tyrosine) was essential for the successful transition of aquatic plants to their vascular dry-land counterparts (Lewis, N.G., and Davin, L.B., in *Isoprenoids and Other Natural Products. Evolution and Function*, 562 (W.D. Nes, ed) 202-246, ACS Symposium Series: Washington, DC (1994)), some four hundred and eighty million years ago (Graham, L.E., *Origin of Land Plants*, John Wiley & Sons, Inc., New York, NY (1993)).

Based on existing chenotaxonomic data, lignans are present in "primitive" plants, such as the fern *Blachium orientale* (Wada, H. et al., *Chem. Pharm. Bull.* 40:2099-2101 (1992)) and the horntworts, e.g., *Dendroceros japonicus* and *Megaceras flagellaris* (Takeda, R. et al., in *Bryophytes. Their Chemistry and Chemical Taxonomy*, Vol. 29 (Zinsmeister, H.D. and Mues, R. eds) pp. 201-207, Oxford University Press: New York, NY (1990); Takeda, R. et al., *Tetrahedron Lett.* 31:4159-4162 (1990)), with the latter recently being classified as originating in the Silurian period (Graham, L.E., *J. Plant Res.* 109: 241-252 (1996)). Interestingly, evolution of both gymnosperms and angiosperms was accompanied by major changes in the structural complexity and oxidative modifications of the lignans (Lewis, N.G., and Davin, L.B., in *Isoprenoids and Other Natural Products. Evolution and Function*, 562 (W. D. Nes, ed) 202-246, ACS Symposium Series: Washington, DC (1994); Gottlieb, O.R., and Yoshida, M., in *Natural Products of Woody Plants. Chemicals Extraneous to the Lignocellulosic Cell Wall* (Rowe, J.W. and Kirk, C.H. eds) pp. 439-511, Springer Verlag: Berlin (1989)). Indeed, in some species, such as Western Red Cedar (*Thuja plicata*), lignans can contribute extensively to heartwood formation/generation by enhancing the resulting heartwood color, quality, fragrance and durability.

In addition to their functions in plants, lignans also have important pharmacological roles. For example, podophyllotoxin, as its etoposide and teniposide derivatives, is an example of a plant compound that has been successfully employed as an anticancer agent (Ayres, D.C., and Loike, J.D. in *Chemistry and Pharmacology of Natural Products. Lignans. Chemical, Biological and Clinical Properties*, Cambridge University Press, Cambridge, England (1990)). Antiviral

properties have also been reported for selected lignans. For example, (γ)-arctigenin (Schröder, H.C. et al., *Z. Naturforsch. 45c*, 1215-1221 (1990)), (γ -trachelogenin (Schröder, H.C. et al., *Z. Naturforsch. 45c*, 1215-1221 (1990)) and nordihydroguaiaretic acid (Gnabre, J.N. et al., *Proc. Natl. Acad. Sci. USA 92*:11239-11243 (1995)) are each effective against HIV due to their pronounced reverse transcriptase inhibitory activities. Some lignans, e.g., matairesinol (Nikaido, T. et al., *Chem. Pharm. Bull.* **29**:3565-3592 (1981)), inhibit cAMP-phosphodiesterase, whereas others enhance cardiovascular activity, e.g., syringaresinol β -D-glucoside (Nishibe, S. et al., *Chem. Pharm. Bull.* **38**:1763-1765 (1990)). There is also a high correlation between the presence, in the diet, of the "mammalian" lignans or "phytoestogens", enterolactone and enterodiol, formed following digestion of high fiber diets, and reduced incidence rates of breast and prostate cancers (so-called chemoprevention) (Axelson, M., and Setchell, K.D.R., *FEBS Lett.* **123**:337-342 (1981); Adlercreutz et al., *J. Steroid Biochem. Molec. Biol.* **41**:3-8 (1992); Adlercreutz et al., *J. Steroid Biochem. Molec. Biol.* **52**:97-103 (1995)). The "mammalian lignans," in turn, are considered to be derived from lignans such as matairesinol and secoisolariciresinol (Borriello et al., *J. Applied Bacteriol.* **58**:37-43 (1985)).

The biosynthetic pathways to the lignans are only now being defined, although there are no prior art reports of the isolation of enzymes or genes involved in the lignan biosynthetic pathway. Based on radiolabeling experiments with crude enzyme extracts from *Forsythia intermedia*, it was first established that entry into the 8,8'-linked lignans, which represent the most prevalent lignol linkage known (Davin, L.B., and Lewis, N.G. in *Rec. Adv. Phytochemistry*, Vol. 26 (Stafford, H.A., and Ibrahim, R.K., eds), pp. 325-375, Plenum Press, New York, NY (1992)), occurs via stereoselective coupling of two achiral coniferyl alcohol molecules, in the form of oxygenated free radicals, to afford the furfuran lignan (+)-pinoresinol (Davin, L.B., Bedgar, D.L., Katayama, T., and Lewis, N.G., *Phytochemistry* **31**:3869-3874 (1992); Paré, P.W. et al., *Tetrahedron Lett.* **35**:4731-4734 (1994)) (FIGURE 1).

30 Bimolecular phenoxy radical coupling reactions, such as the stereoselective coupling of two achiral coniferyl alcohol molecules to afford the furfuran lignan (+)-pinoresinol, are involved in numerous biological processes. These are presumed to include lignin formation in vascular plants (M. Nose et al., *Phytochemistry* **39**:71 (1995)), lignan formation in vascular plants (N.G. Lewis and L.B. Davin, *ACS Symp. Ser.* **562**:202 (1994); P. W. Paré et al., *Tetrahedron Lett.* **35**:4731 (1994)), suberin

formation in vascular plants (M.A. Bernards et al., *J. Biol. Chem.* **270**:7382 (1995)), fruiting body development in fungi (J.D. Bullock et al., *J. Chem. Soc. 2085* (1962)), insect cuticle melanization and sclerotization (M. Miesner et al., *Helv. Chim. Acta 74*:1205 (1991); V.J. Marmaras et al., *Arch. Insect Biochem. Physiol.* **31**:119 (1996)), the formation of aphid pigments (D.W. Cameron and Lord Todd, in *Organic Substances of Natural Origin. Oxidative Coupling of Phenols*, W.I. Taylor and A.R. Battersby, Eds. (Dekker, New York, 1967), Vol. 1, p203), and the formation of algal cell wall polymers (M.A. Ragan, *Phytochemistry* **23**:2029 (1984)).

In contrast to the marked regiochemical and/or stereochemical specificities observed in the biosynthesis of the foregoing lignin and lignan substances *in vivo*, all previously described chemical (I. Iqbal et al., *Chem. Rev.* **94**:519 (1994)) and enzymatic (K. Freudenberg, *Science 148*:595 (1965)) bimolecular phenoxy radical coupling reactions *in vitro* have lacked strict regio- and stereospecific control. That is, if chiral centers are introduced during coupling *in vitro*, the products are racemic, and different regiochemistries can result if more than one potential coupling site is present. Thus, the ability to generate a particular enantiomeric form or a specific coupling product *in vitro* is not under explicit control. Consequently, it is inferred that a mechanism exists *in vivo* to control the regiochemistry and stereochemistry of bimolecular phenoxy radical coupling reactions leading to the formation of, for example, lignans.

In *Forsythia intermedia*, and presumably other species, (+)-pinoresinol, the product of the stereospecific coupling of two E-coniferyl alcohol molecules, undergoes sequential reduction to generate (+)-lanciresinol and then (-)-secoisolariciresinol (Katayama, T. et al., *Phytochemistry* **32**:581-591 (1993); Chu, A. et al., *J. Biol. Chem.* **268**:27026-27033 (1993)) (FIGURE 1). While it has hitherto been unclear whether more than one reductase is required to catalyze the sequential steps, the reductions proceed via abstraction of the pro-R hydride of NADPH, resulting in an "inversion" of configuration at both the C-7 and C-7' positions of the products, (+)-lanciresinol and (-)-secoisolariciresinol (Chu, A., et al., *J. Biol. Chem.* **268**:27026-27033 (1993)). (-)-Matairesinol is subsequently formed via dehydrogenation of (-)-secoisolariciresinol, further metabolism of which presumably affords lignans such as the antiviral (-)-trachelogenin in *Ipomoea carnea* and (-)-podophyllotoxin in *Podophyllum peltatum*.

Thus, the stereospecific formation of (+)-pinoresinol and the subsequent reductive steps giving (+)-lanciresinol and (-)-secoisolariciresinol are pivotal points

in lignan metabolism, since they represent entry into the furano, dibenzylbutane, dibenzylbutyrolactone and aryltetrahydronaphthalene lignan subclasses. Additionally, it should be noted that while lignans are normally optically active, the particular enantiomer present may differ between plant species. For example, (-)-pinoresinol occurs in *Xanthoxylum ailanthoides* (Ishii et al., *Yakugaku Zasshi*, 103:279-292 (1981)), and (-)-larciresinol is present in *Daphne tangutica* (Lin-Gen, et al., *Planta Medica*, 45:172-176 (1982)). The optical activity of a particular lignan may have important ramifications regarding biological activity. For example, (-)-tracheogenin inhibits the *in vitro* replication of HIV-1, whereas its (+)-enantiomer is much less effective (Schroder et al., *Naturforsch. 45c*:1215-1221 (1990)).

Summary of the Invention

In accordance with the foregoing, in one aspect of the invention it has now been discovered that a 78-kD "dirigent" protein is involved in conferring stereospecificity in 8,8'-linked lignan formation. This protein has no detectable catalytically active oxidative center and apparently serves only to bind and orient coniferyl alcohol-derived free radicals, which then undergo stereoselective coupling to form (-)-pinoresinol. The formation of free-radicals, in the first instance, requires the oxidative capacity of either a nonspecific oxidase or even a non-enzymatic electron oxidant. In another aspect of the invention, it has been discovered that a single enzyme, designated pinoresinol/larciresinol reductase, catalyzes the conversion of pinoresinol to larciresinol and then to secoisolarciresinol. Thus, one aspect of the invention relates to isolated "dirigent" proteins and to isolated pinoresinol/larciresinol reductases, such as, for example, those from *Forsythia intermedia*, *Thuya plicata* and *Tsuga heterophylla*.

In other aspects of the invention, cDNAs encoding dirigent protein from *Forsythia intermedia* (SEQ ID Nos:12 and 14), *Thuya plicata* (SEQ ID Nos:20,22,24,26,28,30,32 and 34) and *Tsuga heterophylla* (SEQ ID Nos:16 and 18) have been isolated and sequenced, and the corresponding amino acid sequences have been deduced. Also, cDNAs encoding pinoresinol/larciresinol reductase from *Forsythia intermedia* (SEQ ID Nos:47,49,51,53,55 and 57), *Thuya plicata* (SEQ ID Nos:61,63,65 and 67) and *Tsuga heterophylla* (SEQ ID Nos:69 and 71) have been isolated and sequenced, and the corresponding amino acid sequences have been deduced.

15 In other aspects of the invention, cDNAs encoding dirigent protein from *Forsythia intermedia* (SEQ ID Nos:12 and 14), *Thuya plicata* (SEQ ID Nos:20,22,24,26,28,30,32 and 34) and *Tsuga heterophylla* (SEQ ID Nos:16 and 18) have been isolated and sequenced, and the corresponding amino acid sequences have been deduced. Also, cDNAs encoding pinoresinol/larciresinol reductase from 20 *Forsythia intermedia* (SEQ ID Nos:47,49,51,53,55 and 57), *Thuya plicata* (SEQ ID Nos:61,63,65 and 67) and *Tsuga heterophylla* (SEQ ID Nos:69 and 71) have been isolated and sequenced, and the corresponding amino acid sequences have been deduced.

Thus, the present invention relates to isolated proteins and to isolated DNA sequences which code for the expression of dirigent protein or pinoresinol/larciresinol reductase. In other aspects, the present invention is directed to replicable recombinant cloning vehicles comprising a nucleic acid sequence which codes for a pinoresinol/larciresinol reductase or for a dirigent protein. The present invention is also directed to a base sequence sufficiently complementary to at least a portion of a pinoresinol/larciresinol reductase DNA or RNA, or to at least a portion of a dirigent protein DNA or RNA, to enable hybridization therewith. The aforesaid complementary base sequences include, but are not limited to: antisense pinoresinol/larciresinol reductase RNA; antisense dirigent protein RNA; fragments of DNA that are complementary to a pinoresinol/larciresinol reductase DNA, or to a dirigent protein DNA, and which are therefore useful as polymerase chain reaction primers, or as probes for pinoresinol/larciresinol reductase genes, dirigent protein genes, or related genes.

15 In yet other aspects of the invention, modified host cells are provided that have been transformed, transfected, infected and/or injected with a recombinant cloning vehicle and/or DNA sequence of the invention. Thus, the present invention provides for the recombinant expression of pinoresinol/larciresinol reductases and dirigent proteins in plants, animals, microbes and in cell cultures. The inventive concepts described herein may be used to facilitate the production, isolation and purification of significant quantities of recombinant pinoresinol/larciresinol reductase or dirigent protein, or of their enzyme products, in plants, animals, microbes or cell cultures.

Brief Description of the Drawings

20 The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same becomes better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1 shows the stereospecific conversion of *E*-coniferyl alcohol to (-)-pinoresinol in *Forsythia intermedia*. The stereoselectivity of this reaction is controlled by dirigent protein. (-)-Pinoresinol is then sequentially converted to (-)-larciresinol and (-)-secoisolarciresinol by (-)-pinoresinol/(-)-larciresinol reductase. (-)-pinoresinol, (-)-larciresinol and (-)-secoisolarciresinol are the precursors of the furan, furano and dibenzylbutane families of lignans, respectively.

Detailed Description of the Preferred Embodiment

As used herein, the terms "amino acid" and "amino acids" refer to all naturally occurring L- α -amino acids or their residues. The amino acids are identified by either the single-letter or three-letter designations:

5	Asp	D	aspartic acid	Ile	I	isoleucine
	Thr	T	threonine	Leu	L	leucine
	Ser	S	serine	Tyr	Y	tyrosine
	Glu	E	glutamic acid	Phe	F	phenylalanine
10	Pro	P	proline	His	H	histidine
	Gly	G	glycine	Lys	K	lysine
	Ala	A	alanine	Arg	R	arginine
	Cys	C	cysteine	Trp	W	tryptophan
	Val	V	valine	Gln	Q	glutamine
	Met	M	methionine	Asn	N	asparagine
15						

As used herein, the term "nucleotide" means a monomeric unit of DNA or RNA containing a sugar moiety (pentose), a phosphate and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of pentose) and that combination of base and sugar is called a nucleoside.

The base characterizes the nucleotide with the four bases of DNA being adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). Inosine ("I") is a synthetic base that can be used to substitute for any of the four, naturally-occurring bases (A, C, G or T). The four RNA bases are A, G, C and uracil ("U"). The nucleotide sequences described herein comprise a linear array of nucleotides connected by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

The term "percent identity" (%) means the percentage of amino acids or nucleotides that occupy the same relative position when two amino acid sequences, or two nucleic acid sequences, are aligned side by side.

The term "percent similarity" (%) is a statistical measure of the degree of relatedness of two compared protein sequences. The percent similarity is calculated by a computer program that assigns a numerical value to each compared pair of amino acids based on chemical similarity (e.g., whether the compared amino acids are acidic, basic, hydrophobic, aromatic, etc.) and/or evolutionary distance as measured by the minimum number of base pair changes that would be required to convert a codon encoding one member of a pair of compared amino acids to a codon

encoding the other member of the pair. Calculations are made after a 'best fit' alignment of the two sequences has been made empirically by iterative comparison of all possible alignments. (Henikoff, S. and Henikoff, J.G., *Proc. Natl' Acad. Sci. USA* 89:10915-10919 (1992)).

5 "Oligonucleotide" refers to short length single or double stranded sequences of deoxyribonucleotides linked via phosphodiester bonds. The oligonucleotides are chemically synthesized by known methods and purified, for example, on polyacrylamide gels.

The term "pinoresinol/laricresinol reductase" is used herein to mean an enzyme capable of catalyzing two reduction reactions: the reduction of pinoresinol to laricresinol, and the reduction of laricresinol to secoisolaricresinol. The products of these reactions, laricresinol and secoisolaricresinol, can be either the (+)- or (-)-enantiomers.

The term "dirigent protein" is used herein to mean a protein capable of guiding a bimolecular phenoxy radical coupling reaction thereby determining the stereochemistry and regiochemistry of the product of the reaction and/or its polymeric derivatives.

The terms "alteration", "amino acid sequence alteration", "variant" and "amino acid sequence variant" refer to dirigent protein or pinoresinol/laricresinol reductase molecules with some differences in their amino acid sequences as compared to the corresponding native dirigent protein or pinoresinol/laricresinol reductase. Ordinarily, the variants will possess at least about 70% homology with the corresponding, native dirigent protein or pinoresinol/laricresinol reductase, and preferably they will be at least about 80% homologous with the corresponding, native

20 variants of dirigent protein or pinoresinol/laricresinol reductase falling within this invention possess substitutions, deletions, and/or insertions at certain positions. Sequence variants of dirigent protein or pinoresinol/laricresinol reductase may be used to attain desired enhanced or reduced enzymatic activity, modified regiochemistry or stereochemistry, or altered substrate utilization or product distribution.

Substitutional dirigent protein variants or pinoresinol/laricresinol reductase variants are those that have at least one amino acid residue in the corresponding native dirigent protein sequence or pinoresinol/laricresinol reductase sequence removed and a different amino acid inserted in its place at the same position. The

35

substitutions may be single, where only one amino acid in the molecule has been substituted, or they may be multiple, where two or more amino acids have been substituted in the same molecule. Substantial changes in the activity of the dirigent protein or pinoresinol/lariciresinol reductase molecule may be obtained by substituting an amino acid with a side chain that is significantly different in charge and/or structure from that of the native amino acid. This type of substitution would be expected to affect the structure of the polypeptide backbone and/or the charge or hydrophobicity of the molecule in the area of the substitution.

Moderate changes in the activity of the dirigent protein or pinoresinol/lariciresinol reductase molecule would be expected by substituting an amino acid with a side chain that is similar in charge and/or structure to that of the native molecule. This type of substitution, referred to as a conservative substitution, would not be expected to substantially alter either the structure of the polypeptide backbone or the charge or hydrophobicity of the molecule in the area of the substitution.

Insertional dirigent protein variants or pinoresinol/lariciresinol reductase variants are those with one or more amino acids inserted immediately adjacent to an amino acid at a particular position in the native dirigent protein or pinoresinol/lariciresinol reductase molecule. Immediately adjacent to an amino acid means connected to either the α -carboxy or α -amino functional group of the amino acid.

The insertion may be one or more amino acids. Ordinarily, the insertion will consist of one or two conservative amino acids. Amino acids similar in charge and/or structure to the amino acids adjacent to the site of insertion are defined as conservative. Alternatively, this invention includes insertion of an amino acid with a charge and/or structure that is substantially different from the amino acids adjacent to the site of insertion.

Deletional variants are those where one or more amino acids in the native dirigent protein or pinoresinol/lariciresinol reductase molecule have been removed. Ordinarily, deletional variants will have one or two amino acids deleted in a particular region of the dirigent protein or pinoresinol/lariciresinol reductase molecule.

The term "antisense" or "antisense RNA" or "antisense nucleic acid" is used herein to mean a nucleic acid molecule that is complementary to all or part of a messenger RNA molecule. Antisense nucleic acid molecules are typically used to inhibit the expression, *in vivo*, of complementary, expressed messenger RNA molecules.

The terms "biological activity", "biologically active", "activity" and "active" when used with reference to a pinoresinol/lariciresinol reductase molecule refer to the ability of the pinoresinol/lariciresinol reductase molecule to reduce pinoresinol and lariciresinol to yield lariciresinol and secosolariresinol, respectively, as measured in an enzyme activity assay, such as the assay described in Example 8 below.

The terms "biological activity", "biologically active", "activity" and "active" when used with reference to a dirigent protein refer to the ability of the dirigent protein to guide a bimolecular phenoxy radical coupling reaction thereby determining the stereochemistry and regiochemistry of the product of the reaction and of its polymeric derivatives.

Amino acid sequence variants of dirigent protein or pinoresinol/lariciresinol reductase may have desirable altered biological activity including, for example, altered reaction kinetics, substrate utilization, product distribution or other characteristics such as regiochemistry and stereochemistry.

The terms "DNA sequence encoding", "DNA encoding" and "nucleic acid encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the translated polypeptide chain. The DNA sequence thus codes for the amino acid sequence.

The terms "replicable expression vector" and "expression vector" refer to a piece of DNA, usually double-stranded, which may have inserted into it a piece of foreign DNA. Foreign DNA is defined as heterologous DNA, which is DNA not naturally found in the host. The vector is used to transport the foreign or heterologous DNA into a suitable host cell. Once in the host cell, the vector can replicate independently of or coincidentally with the host chromosomal DNA, and several copies of the vector and its inserted (foreign) DNA may be generated. In addition, the vector contains the necessary elements that permit translating the foreign DNA into a polypeptide. Many molecules of the polypeptide encoded by the foreign DNA can thus be rapidly synthesized.

The terms "transformed host cell", "transformed" and "transformation" refer to the introduction of DNA into a cell. The cell is termed a "host cell", and it may be a prokaryotic or a eukaryotic cell. Typical prokaryotic host cells include various strains of *E. coli*. Typical eukaryotic host cells are plant cells, such as maize cells, yeast cells, insect cells or animal cells. The introduced DNA is usually in the form of

a vector containing an inserted piece of DNA. The introduced DNA sequence may be from the same species as the host cell or from a different species from the host cell, or it may be a hybrid DNA sequence, containing some foreign DNA and some DNA derived from the host species.

In accordance with the present invention, cDNAs encoding dirigent protein and pinoresinol/lariciresinol reductase from *Forsythia intermedia*, *Thuya plicata* and *Tsuga heterophylla* were isolated, sequenced and expressed in the following manner. With respect to the cDNAs encoding dirigent protein from *Forsythia intermedia*, an empirically-determined purification protocol was developed to isolate the *Forsythia* dirigent protein. This procedure yielded at least six isoforms of the dirigent protein. Amino acid sequencing of the amino terminus of each of these isoforms revealed that the sequence of each isoform was identical. Sequencing of the N-terminus of a mixture of these isoforms yielded a 28 amino acid sequence (SEQ ID No:1). Tryptic digestion of a mixture of these isoforms yielded six peptide fragments which were purified in sufficient quantity to permit sequencing (SEQ ID Nos:2-7).

A primer designated PSINT1 (SEQ ID No:8) was synthesized based on the sequence of amino acids 9 to 15 of the N-terminal peptide (SEQ ID No:1). A primer designated PS1R (SEQ ID No:9) was synthesized based on the sequence of amino acids 3 to 9 of the internal peptide sequence set forth in (SEQ ID No:2). A primer designated PS2R (SEQ ID No:10) was synthesized based on the sequence of amino acids 13 to 20 of the internal peptide sequence set forth in (SEQ ID No:2). A primer designated PS3R (SEQ ID No:11) was synthesized based on the sequence of amino acids 6 to 12 of the internal peptide sequence set forth in (SEQ ID No:3).

Forsythia total RNA was isolated by means of a protocol adapted from a method specifically designed for woody tissues which contain a large concentration of polyphenols. Poly A+ RNA was isolated and a cDNA library constructed using standard means. A PCR reaction utilizing primers PSINT1 (SEQ ID No:8) and one of PS17R, (SEQ ID No:11) PS12R (SEQ ID No:10) or PS1R (SEQ ID No:9), together with an aliquot of *Forsythia* cDNA as substrate, each yielded a single cDNA band of ~370 bp, ~155 bp and ~125 bp, respectively. The ~370 bp product of the PSINT1 (SEQ ID NO:8)-PS17R (SEQ ID No:11) reaction was amplified by PCR and utilized as a probe to screen approximately 600,000 PFU of a *Forsythia intermedia* cDNA library. Two distinct cDNAs were identified, called pPSDF1 (SEQ ID No:12) and pPSDF2 (SEQ ID No:14). The cDNA insert encoding dirigent

protein was excised from plasmid pPSDF1 and cloned into the baculovirus transfer vector pBlueBact. The resulting construct was used to transform *Spodoptera frugiperda* from which functional dirigent protein was purified.

With respect to the cloning of dirigent protein from *Thuya plicata* and *Tsuga heterophylla*, the *Forsythia* cDNAs were used as probes to isolate two dirigent protein clones from *Tsuga heterophylla* (SEQ ID Nos:16, 18), and eight dirigent protein cDNA clones from *Thuya plicata* (SEQ ID Nos:20, 22, 24, 26, 28, 30, 32, 34).

With respect to the cDNAs encoding (+)-pinoresinol/(-)-lariciresinol reductase from *Forsythia intermedia*, an empirically-determined purification protocol, consisting of eight chromatographic steps, was developed to isolate the *Forsythia* (+)-pinoresinol/(-)-lariciresinol reductase protein. This procedure yielded two isoforms of (+)-pinoresinol/(-)-lariciresinol reductase which were both capable of catalyzing the reduction of (+)-pinoresinol and (-)-lariciresinol. Sequencing of the N-terminus of each of these isoforms yielded an identical 30 amino acid sequence (SEQ ID No:36). Tryptic digestion of a mixture of both of these isoforms yielded four peptide fragments which were purified in sufficient quantity to permit sequencing (SEQ ID Nos:37-40). Additionally, cyanogen bromide cleavage of a mixture of both of these isoforms yielded three peptide fragments which were purified in sufficient quantity to permit sequencing (SEQ ID Nos:41-43).

A primer designated PLRNS (SEQ ID No:44) was synthesized based on the sequence of amino acids 7 to 13 of the N-terminal peptide (SEQ ID No:36). A primer designated PLR14R (SEQ ID No:45) was synthesized based on the sequence of amino acids 2 to 8 of the internal peptide sequence set forth in SEQ ID No:37. A primer designated PLR15R (SEQ ID No:46) was synthesized based on the sequence of amino acids 9 to 15 of the internal peptide sequence set forth in SEQ ID No:45.

The sequence of amino acids 9 to 15 of the internal peptide sequence set forth in SEQ ID No:37, upon which the sequence of primer PLR15R (SEQ ID No:46) was based, also corresponded to the sequence of amino acids 4 to 10 of the cyanogen bromide-generated, internal fragment RNA was isolated by means of a protocol adapted from a

Forsythia total RNA was isolated by means of a protocol adapted from a method specifically designed for woody tissues which contain a large concentration of polyphenols. Poly A+ RNA was isolated and a cDNA library constructed using standard means. A PCR reaction utilizing primers PLRNS (SEQ ID No:44) and either PLR14R (SEQ ID No:45) or PLR15R (SEQ ID No:46), together with an

aliquot of *Forsythia* cDNA as substrate, yielded two, amplified bands of 380 bp and 400 bp. One 400 bp cDNA insert was utilized as a probe with which to screen the *Forsythia* cDNA library. The 400 bp probe corresponded to bases 22 to 423 of SEQ ID No:47. Six cDNA clones were isolated and sequenced (SEQ ID Nos:47; 49; 51; 53; 55; 57). The clones shared a common coding region, many had a different 5' untranslated region and the 3'-untranslated region of each terminated at a different point. One of these cDNAs (SEQ ID No:47), expressed as a β -galactosidase fusion protein in *E. coli*, catalyzed the same enantiomer-specific reactions as the native plant protein.

With respect to the cloning of (+)-pinoresinol/(-)-laricresinol reductase and (-)-pinoresinol/(-)-laricresinol reductase from *Trigia pilifera*, cDNA was synthesized and utilized as a template in a PCR reaction in which the primers were a 3' linker-primer (SEQ ID No:59) and a 5' primer, designated CR6-NT, (SEQ ID No:60). At least two bands of the expected length (1.2 kb) were generated and cloned into a 10 plasmid vector. One clone, designated pIr-Tp1, (SEQ ID No:61) was completely sequenced and expressed as a β -galactosidase fusion protein in *E. coli*. pIr-Tp1 encodes a (+)-pinoresinol/(-)-laricresinol reductase.

The cDNA insert of clone pIr-Tp1 was used to screen the *T. pilifera* cDNA library and identified an additional, unique clone, designated pIr-Tp2, (SEQ ID No:63). pIr-Tp2 has high homology to pIr-Tp1 but encodes a (+)-pinoresinol/(-)-laricresinol reductase. The cDNA insert of clone pIr-Tp1 was used to screen the *T. pilifera* cDNA library and identify an additional two pinoresinol/laricresinol reductase cDNAs (SEQ ID Nos:65; 67).

Two cDNAs encoding pinoresinol/laricresinol reductases from *Trigia heterophylla* (SEQ ID Nos:69; 71) were isolated by screening a *Trigia heterophylla* cDNA library with the pIr-Tp1 cDNA insert. The isolation of cDNAs encoding dirigent proteins, (+)-pinoresinol/(-)-laricresinol reductase permits the development of an efficient expression system for these functional enzymes; 25 provides useful tools for examining the developmental regulation of lignan biosynthesis and permits the isolation of other dirigent proteins and pinoresinol/laricresinol reductases. The isolation of the dirigent protein and pinoresinol/laricresinol reductase cDNAs also permits the transformation of a wide range of organisms in order to enhance or modify lignan biosynthesis.

The proteins and nucleic acids of the present invention can be utilized to predetermine the stereochemistry, regiochemistry, or both, of the products of bimolecular phenoxy coupling reactions, such as the furofuran, furano and dibenzylbutane lignans. By way of nonlimiting examples, the proteins and nucleic acids of the present invention can be utilized to: elevate or otherwise alter the levels of health-protecting lignans, such as podophyllotoxin, in plant species, including but not limited to vegetables, grains and fruits, and to food items incorporating material derived from such genetically altered plants; genetically alter plant species to provide an abundant, natural supply of lignans useful for a variety of purposes, for example as neutraceuticals and dietary supplements; to genetically alter living organisms to produce an abundant supply of optically pure lignans having desirable biological properties, for example (-)-arctigenin which possesses antiviral properties. In particular, characterization of the dirigent protein binding site and mechanism of action permits the development of synthetic proteins consisting of an array of 5 dirigent protein binding sites which serve as templates for stereochemically-controlled polymeric assembly.

N-terminal transport sequences well known in the art (see, e.g., von Heijne, G. et al, *Eur. J. Biochem.* 180:535-545 (1989); Stryer, *Biochemistry* W.H. Freeman and Company, New York, NY, p. 769 (1988)) may be employed to direct the dirigent protein or pinoresinol/laricresinol reductase to a variety of cellular or extracellular locations.

Sequence variants of wild-type dirigent protein clones and pinoresinol/laricresinol clones that can be produced by deletions, substitutions, mutations and/or insertions are intended to be within the scope of the invention except insofar as limited by the prior art. Dirigent protein or pinoresinol/laricresinol reductase amino acid sequence variants may be constructed by mutating the DNA sequence that encodes wild-type dirigent protein or wild-type pinoresinol/laricresinol reductase, such as by using techniques commonly referred to as site-directed mutagenesis. Various polymerase chain reaction (PCR) methods now well known in the field, such as a two primer system like the Transformer Site-Directed Mutagenesis kit from 10 Clontech, may be employed for this purpose.

Following denaturation of the target plasmid in this system, two primers are simultaneously annealed to the plasmid; one of these primers contains the desired site-directed mutation, the other contains a mutation at another point in the plasmid 15 resulting in elimination of a restriction site. Second strand synthesis is then carried

35

-15-

out, tightly linking these two mutations, and the resulting plasmids are transformed into a *muS* strain of *E. coli*. Plasmid DNA is isolated from the transformed bacteria, restricted with the relevant restriction enzyme (thereby linearizing the unmutated plasmids), and then retransformed into *E. coli*. This system allows for generation of mutations directly in an expression plasmid, without the necessity of subcloning or generation of single-stranded phagemids. The tight linkage of the two mutations and the subsequent linearization of unmutated plasmids results in high mutation efficiency and allows minimal screening. Following synthesis of the initial restriction site primer, this method requires the use of only one new primer type per mutation site. Rather than prepare each positional mutant separately, a set of "designed degenerate" oligonucleotide primers can be synthesized in order to introduce all of the desired mutations at a given site simultaneously. Transformants can be screened by sequencing the plasmid DNA through the mutagenized region to identify and sort mutant clones. Each mutant DNA can then be restricted and analyzed by electrophoresis on Mutation Detection Enhancement gel (J.T. Baker) to confirm that no other alterations in the sequence have occurred (by band shift comparison to the unmutagenized control).

The verified mutant duplexes can be cloned into a replicable expression vector, if not already cloned into a vector of this type, and the resulting expression construct used to transform *E. coli*, such as strain *E. coli* BL21(DE3)pLysS, for high level production of the mutant protein, and subsequent purification thereof. The method of FAB-MS mapping can be employed to rapidly check the fidelity of mutant expression. This technique provides for sequencing segments throughout the whole protein and provides the necessary confidence in the sequence assignment. In a mapping experiment of this type, protein is digested with a protease (the choice will depend on the specific region to be modified since this segment is of prime interest and the remaining map should be identical to the map of unmutagenized protein). The set of cleavage fragments is fractionated by microbore HPLC (reversed phase or ion exchange, again depending on the specific region to be modified) to provide several peptides in each fraction, and the molecular weights of the peptides are determined by FAB-MS. The masses are then compared to the molecular weights of peptides expected from the digestion of the predicted sequence, and the correctness of the sequence quickly ascertained. Since this mutagenesis approach to protein modification is directed, sequencing of the altered peptide should not be necessary if the MS agrees with prediction. If necessary to verify a changed residue,

CAD-tandem MS/MS can be employed to sequence the peptides of the mixture in question, or the target peptide purified for subtractive Edman degradation or carboxypeptidase Y digestion depending on the location of the modification.

In the design of a particular site directed mutant, it is generally desirable to first make a non-conservative substitution (e.g., Ala for Cys, His or Glu) and determine if activity is greatly impaired as a consequence. The properties of the mutagenized protein are then examined with particular attention to the kinetic parameters of K_m and k_{cat} as sensitive indicators of altered function, from which changes in binding and/or catalysis *per se* may be deduced by comparison to the native enzyme. If the residue is by this means demonstrated to be important by activity impairment, or knockout, then conservative substitutions can be made, such as Asp for Glu to alter side chain length, Ser for Cys, or Arg for His. For hydrophobic segments, it is largely size that will be altered, although aromatics can also be substituted for alkyl side chains. Changes in the normal product distribution can indicate which step(s) of the reaction sequence have been altered by the mutation.

Other site directed mutagenesis techniques may also be employed with the nucleotide sequences of the invention. For example, restriction endonuclease digestion of DNA followed by ligation may be used to generate dirigent protein or 20 pinoresinol/lariciresinol reductase deletion variants, as described in Section 15.3 of Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, New York, NY (1989)). A similar strategy may be used to construct insertion variants, as described in Section 15.3 of Sambrook et al, *supra*.

Oligonucleotide-directed mutagenesis may also be employed for preparing 25 substitution variants of this invention. It may also be used to conveniently prepare the deletion and insertion variants of this invention. This technique is well known in the art as described by Adelman et al. (*DNA* 2:183 (1983)). Generally, oligonucleotides of at least 25 nucleotides in length are used to insert, delete or substitute two or more nucleotides in the dirigent protein gene or pinoresinol/lariciresinol reductase gene. An optimal oligonucleotide will have 12 to 15 perfectly matched nucleotides on either side of the nucleotides coding for the mutation. To 30 mutagenize the wild-type dirigent protein or wild-type pinoresinol/lariciresinol reductase, the oligonucleotide is annealed to the single-stranded DNA template molecule under suitable hybridization conditions. A DNA polymerizing enzyme, usually the Klenow fragment of *E. coli* DNA polymerase I, is then added. This

-16-

enzyme uses the oligonucleotide as a primer to complete the synthesis of the mutation-bearing strand of DNA. Thus, a heteroduplex molecule is formed such that one strand of DNA encodes the wild-type dirigent protein or pinoresinol/lariciresinol reductase inserted in the vector, and the second strand of DNA encodes the mutated form of dirigent protein or pinoresinol/lariciresinol reductase inserted into the same vector. This heteroduplex molecule is then transformed into a suitable host cell.

Mutants with more than one amino acid substituted may be generated in one of several ways. If the amino acids are located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If, however, the amino acids are located some distance from each other (separated by more than ten amino acids, for example) it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed. In the first method, a separate oligonucleotide is generated for each amino acid to be substituted.

The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions.

An alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants. 20 wild-type dirigent protein or pinoresinol/lariciresinol reductase DNA is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or 25 more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third round of mutagenesis, and so on.

Eukaryotic expression systems may be utilized for dirigent protein or pinoresinol/lariciresinol reductase production since they are capable of carrying out any required posttranslational modifications and of directing the enzyme to the proper membrane location. A representative eukaryotic expression system for this purpose uses the recombinant baculovirus, *Autographa californica* nuclear polyhedrosis virus (AcNPV; M.D. Summers and G.E. Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures* (1986); Luckow et al.,

Bio-technology 6:47-55 (1987)) for expression of the dirigent protein or pinoresinol/lariciresinol reductases of the invention. Infection of insect cells (such as cells of the species *Spodoptera frugiperda*) with the recombinant baculoviruses allows for the production of large amounts of the dirigent protein or pinoresinol/lariciresinol reductase protein. In addition, the baculovirus system has other important advantages for the production of recombinant dirigent protein or pinoresinol/lariciresinol reductase. For example, baculoviruses do not infect humans and can therefore be safely handled in large quantities. In the baculovirus system, a DNA construct is prepared including a DNA segment encoding dirigent protein or pinoresinol/lariciresinol reductase and a vector. The vector may comprise the polyhedron gene promoter region of a baculovirus, the baculovirus flanking sequences necessary for proper cross-over during recombination (the flanking sequences comprise about 200-300 base pairs adjacent to the promoter sequence) and a bacterial origin of replication which permits the construct to replicate in bacteria.

15 The vector is constructed so that (i) the DNA segment is placed adjacent (or operably-linked or "downstream" or "under the control of") to the polyhedron gene promoter and (ii) the promoter/pinoresinol/lariciresinol reductase, or promoter/-dirigent protein, combination is flanked on both sides by 200-300 base pairs of baculovirus DNA (the flanking sequences).

20 To produce a dirigent protein DNA construct, or a pinoresinol/lariciresinol reductase DNA construct, a cDNA clone encoding a full length dirigent protein or pinoresinol/lariciresinol reductase is obtained using methods such as those described herein. The DNA construct is contacted in a host cell with baculovirus DNA of an appropriate baculovirus (that is, of the same species of baculovirus as the promoter encoded in the construct) under conditions such that recombination is effected. The resulting recombinant baculoviruses encode the full "dirigent" protein or pinoresinol/lariciresinol reductase. For example, an insect host cell can be cotransfected or transfected separately with the DNA construct and a functional baculovirus. Resulting recombinant baculoviruses can then be isolated and used to infect cells to effect production of dirigent protein or pinoresinol/lariciresinol reductase. Host insect cells include, for example, *Spodoptera frugiperda* cells. Insect host cells infected with a recombinant baculovirus of the present invention are then cultured under conditions allowing expression of the baculovirus-encoded dirigent protein or pinoresinol/lariciresinol reductase. Recombinant protein thus produced is then extracted from the cells using methods known in the art.

Other eukaryotic microbes such as yeasts may also be used to practice this invention. The baker's yeast *Saccharomyces cerevisiae*, is a commonly used yeast, although several other strains are available. The plasmid YRp7 (Stinchcomb et al., *Nature* 282:59 (1979); Kingsman et al., *Gene* 7:141 (1979); Tschemper et al., *Gene* 10:157 (1980)) is commonly used as an expression vector in *Saccharomyces*. This plasmid contains the *trp1* gene that provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, such as strains ATCC No. 44,076 and PEP4-1 (Jones, *Genetics* 85:12 (1977)). The presence of the *trp1* lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan. Yeast host cells are generally transformed using the polyethylene glycol method, as described by Hinnen (Proc. Natl. Acad. Sci. USA 75:1929 (1978)). Additional yeast transformation protocols are set forth in Gietz et al., *N.A.R.* 20(17):1425 (1992); Reeves et al., *FEMS* 99:191-197 (1992).

15 Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem.* 255:2073 (1980)) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg.* 7:149 (1968); Holland et al., *Biochemistry* 17:4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In the construction of suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination. Other promoters that have the additional advantage of transcription controlled by growth conditions are the promoter region for alcohol dehydrogenase 2, isocytchrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Any plasmid vector containing 20 yeast-compatible promoter, origin of replication and termination sequences is suitable.

25 Cell cultures derived from multicellular organisms, such as plants, may be used as hosts to practice this invention. Transgenic plants can be obtained, for example, by transferring plasmids that encode pinoresinol/kaurene reductase, and/or dirigent protein, and a selectable marker gene, e.g., the kan gene encoding 30

resistance to kanamycin, into *Agrobacterium tumefaciens* containing a helper Ti plasmid as described in Hooekema et al., *Nature* 303:179-181 (1983) and culturing the *Agrobacterium* cells with leaf slices of the plant to be transformed as described by An et al., *Plant Physiology* 81:301-305 (1986). Transformation of cultured plant host cells is normally accomplished through *Agrobacterium tumefaciens*, as described above. Cultures of mammalian host cells and other host cells that do not have rigid cell membrane barriers are usually transformed using the calcium phosphate method as originally described by Graham and Van der Eb (*Virology* 52:546 (1978)) and modified as described in Sections 16.32-16.37 of Sambrook et al., *supra*. However, other methods for introducing DNA into cells such as Polybrene (Kawai and Nishizawa, *Mol. Cell. Bio.* 4:1172 (1984)), protoplast fusion (Schaffner, *Proc. Natl. Acad. Sci. USA* 77:2163 (1980)), electroporation (Neumann et al., *EMBO J.* 1:841 (1982)), and direct microinjection into nuclei (Capecchi, *Cell* 22:479 (1980)) may also be used. Additionally, animal transformation strategies are reviewed in 15 Monastersky G.M. and Robl, J.M., *Strategies in Transgenic Animal Science*, ASM Press, Washington, D.C. (1995). Transformed plant calli may be selected through the selectable marker by growing the cells on a medium containing, e.g., kanamycin, and appropriate amounts of phytohormone such as naphthalene acetic acid and benzyladenine for callus and shoot induction. The plant cells may then be regenerated and the resulting plants transferred to soil using techniques well known to those skilled in the art.

In addition, a gene regulating pinoresinol/kaurene reductase production, or dirigent protein production, can be incorporated into the plant along with a necessary promoter which is inducible. In the practice of this embodiment of the invention, a promoter that only responds to a specific external or internal stimulus is fused to the target cDNA. Thus, the gene will not be transcribed except in response to the specific stimulus. As long as the gene is not being transcribed, its gene product is not produced.

20 An illustrative example of a responsive promoter system that can be used in the practice of this invention is the glutathione-S-transferase (GST) system in maize. GSTs are a family of enzymes that can detoxify a number of hydrophobic electrophilic compounds that often are used as pre-emergent herbicides (Weigand et al., *Plant Molecular Biology* 7:235-243 (1986)). Studies have shown that the GSTs are directly involved in causing this enhanced herbicide tolerance. 25 This action is primarily mediated through a specific 1.1 kb mRNA transcription

product. In short, maize has a naturally occurring quiescent gene already present that can respond to external stimuli and that can be induced to produce a gene product. This gene has previously been identified and cloned. Thus, in one embodiment of this invention, the promoter is removed from the GST responsive gene and attached to a pinoresinol/antiresinol reductase gene, or a dirigent protein gene, that previously has had its native promoter removed. This engineered gene is the combination of a promoter that responds to an external chemical stimulus and a gene responsible for successful production of pinoresinol/antiresinol reductase or dirigent protein.

In addition to the methods described above, several methods are known in the art for transferring cloned DNA into a wide variety of plant species, including gymnosperms, angiosperms, monocots and dicots (see, e.g., Glick and Thompson, eds., *Methods in Plant Molecular Biology*, CRC Press, Boca Raton, Florida (1993)). Representative examples include electroporation-facilitated DNA uptake by protoplasts (Rhodes et al., *Science* 240(4849):204-207 (1988); treatment of protoplasts with polyethylene glycol (Uyznik et al., *Plant Molecular Biology* 13:151-161 (1989)); and bombardment of cells with DNA laden microparticles (Klein et al., *Plant Physiol.* 91:440-444 (1989) and Boynton et al., *Science* 240(4858):1534-1538 (1988)). Numerous methods now exist, for example, for the transformation of cereal crops (see, e.g., McKinnon, G.E. and Henry, R.T. *J. Cereal Science*, 22(3):203-210 (1995); Mendel, R.R. and Teeri, T.H. *Plant and Microbial Biotechnology Research Series*, 3:81-98, Cambridge University Press (1995); McElroy, D. and Brettell, R.S., *Trends in Biotechnology* 12(2):52-68 (1994); Christou et al., *Trends in Biotechnology* 10(7):239-246 (1992); Christou, P. and Ford, T.L., *Annals of Botany*, 75(5): 449-454 (1995); Park et al., *Plant Molecular Biology*, 32(6):1135-1148 (1996); Altpeter et al., *Plant Cell Reports*, 16:12-17 (1996)). Additionally, plant transformation strategies and techniques are reviewed in Birch, R.G., *Annu Rev Plant Phys Plant Mol Biol* 48:297 (1997); Forester et al., *Exp. Agric.* 33:15-33 (1997).

Minor variations make these technologies applicable to a broad range of plant species.

Each of these techniques has advantages and disadvantages. In each of the techniques, DNA from a plasmid is genetically engineered such that it contains not only the gene of interest, but also selectable and screenable marker genes. A selectable marker gene is used to select only those cells that have integrated copies of the plasmid (the construction is such that the gene of interest and the selectable and

screenable genes are transferred as a unit). The screenable gene provides another check for the successful culturing of only those cells carrying the genes of interest. A commonly used selectable marker gene is neomycin phosphotransferase II (NPT II). This gene conveys resistance to kanamycin, a compound that can be added directly to the growth media on which the cells grow. Plant cells are normally susceptible to kanamycin and, as a result, die. The presence of the NPT II gene overcomes the effects of the kanamycin and each cell with this gene remains viable. Another selectable marker gene which can be employed in the practice of this invention is the gene which confers resistance to the herbicide glufosinate (Basta). A screenable gene characterized using a histochemical reaction in which a sample of putatively transformed cells is treated with a GUS assay solution. After an appropriate incubation, the cells containing the GUS gene turn blue. Preferably, the plasmid will contain both selectable and screenable marker genes.

The plasmid containing one or more of these genes is introduced into either plant protoplasts or callus cells by any of the previously mentioned techniques. If the marker gene is a selectable gene, only those cells that have incorporated the DNA package survive under selection with the appropriate phytotoxic agent. Once the appropriate cells are identified and propagated, plants are regenerated. Progeny from the transformed plants must be tested to insure that the DNA package has been successfully integrated into the plant genome.

Mammalian host cells may also be used in the practice of the invention. Examples of suitable mammalian cell lines include monkey kidney CV1 line transformed by SV40 (CCS-7, ATCC CRL 1651); human embryonic kidney line 293S (Graham et al., *J. Gen. Virol.* 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells (Ulab and Chasin, *Proc. Natl. Acad. Sci. USA* 77:1216 (1980)); mouse teratocarcinoma cells (TM4, Mather, *Biol. Reprod.* 23:243 (1980)); monkey kidney cells (CV1-76, ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, BRL 3A, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 10, ATCC CCL 2); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor cells (MMT 060562, ATCC CCL 51); rat hepatoma cells (HTC, ML54, Baumann et al., *J. Cell. Biol.* 85:1 (1980)); and TR1 cells (Mather et al., *Annals N.Y. Acad. Sci.* 393:44 (1982)).

Expression vectors for these cells ordinarily include (if necessary) DNA sequences

5 commonly used is the β -glucuronidase gene (GUS). The presence of this gene is characterized using a histochemical reaction in which a sample of putatively transformed cells is treated with a GUS assay solution. After an appropriate incubation, the cells containing the GUS gene turn blue. Preferably, the plasmid will

10 contain both selectable and screenable marker genes.

15 The plasmid containing one or more of these genes is introduced into either plant protoplasts or callus cells by any of the previously mentioned techniques. If the marker gene is a selectable gene, only those cells that have incorporated the DNA package survive under selection with the appropriate phytotoxic agent. Once the appropriate cells are identified and propagated, plants are regenerated. Progeny from the transformed plants must be tested to insure that the DNA package has been

20 successfully integrated into the plant genome.

Mammalian host cells may also be used in the practice of the invention. Examples of suitable mammalian cell lines include monkey kidney CV1 line transformed by SV40 (CCS-7, ATCC CRL 1651); human embryonic kidney line 293S (Graham et al., *J. Gen. Virol.* 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells (Ulab and Chasin, *Proc. Natl. Acad. Sci. USA* 77:1216 (1980)); mouse teratocarcinoma cells (TM4, Mather, *Biol. Reprod.* 23:243 (1980)); monkey kidney cells (CV1-76, ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, BRL 3A, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 10, ATCC CCL 2); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor cells (MMT 060562, ATCC CCL 51); rat hepatoma cells (HTC, ML54, Baumann et al., *J. Cell. Biol.* 85:1 (1980)); and TR1 cells (Mather et al., *Annals N.Y. Acad. Sci.* 393:44 (1982)).

25 Expression vectors for these cells ordinarily include (if necessary) DNA sequences

30

35

for an origin of replication, a promoter located in front of the gene to be expressed, a ribosome binding site, an RNA splice site, a polyadenylation site, and a transcription terminator site.

Promoters used in mammalian expression vectors are often of viral origin. These viral promoters are commonly derived from polyoma virus, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The SV40 virus contains two promoters that are termed the early and late promoters. These promoters are particularly useful because they are both easily obtained from the virus as one DNA fragment that also contains the viral origin of replication (Fiers et al., *Nature* 273:113 (1978)). Smaller or larger SV40 DNA fragments may also be used, provided they contain the approximately 250-bp sequence extending from the HindIII site toward the BglI site located in the viral origin of replication.

Alternatively, promoters that are naturally associated with the foreign gene (homologous promoters) may be used provided that they are compatible with the host cell line selected for transformation.

An origin of replication may be obtained from an exogenous source, such as SV40 or other virus (e.g., Polyoma, Adeno, VSV, BPV) and inserted into the cloning vector. Alternatively, the origin of replication may be provided by the host cell chromosomal replication mechanism. If the vector containing the foreign gene is integrated into the host cell chromosome, the latter is often sufficient.

The use of a secondary DNA coding sequence can enhance production levels of pinoresinol/lariciresinol reductase or dirigent protein in transformed cell lines. The secondary coding sequence typically comprises the enzyme dihydrofolate reductase (DHFR). The wild-type form of DHFR is normally inhibited by the chemical methotrexate (MTX). The level of DHFR expression in a cell will vary depending on the amount of MTX added to the cultured host cells. An additional feature of DHFR that makes it particularly useful as a secondary sequence is that it can be used as a selection marker to identify transformed cells. Two forms of DHFR are available for use as secondary sequences, wild-type DHFR and MTX-resistant DHFR. The type of DHFR used in a particular host cell depends on whether the host cell is DHFR deficient (such that it either produces very low levels of DHFR endogenously, or it does not produce functional DHFR at all). DHFR-deficient cell lines such as the CHO cell line described by Urlaub and Chasin, *supra*, are transformed with wild-type DHFR coding sequences. After transformation, these DHFR-deficient cell lines express functional DHFR and are capable of growing in a

culture medium lacking the nutrients hypoxanthine, glycine and thymidine. Nontransformed cells will not survive in this medium.

The MTX-resistant form of DHFR can be used as a means of selecting for transformed host cells in those host cells that endogenously produce normal amounts of functional DHFR that is MTX sensitive. The CHO-K1 cell line (ATCC No. CL 61) possesses these characteristics, and is thus a useful cell line for this purpose. The addition of MTX to the cell culture medium will permit only those cells transformed with the DNA encoding the MTX-resistant DHFR to grow. The nontransformed cells will be unable to survive in this medium.

Prokaryotes may also be used as host cells for the initial cloning steps of this invention. They are particularly useful for rapid production of large amounts of DNA, for production of single-stranded DNA templates used for site-directed mutagenesis, for screening many mutants simultaneously, and for DNA sequencing of the mutants generated. Suitable prokaryotic host cells include *E. coli* K12 strain 10294 (ATCC No. 31,446), *E. coli* strain W3110 (ATCC No. 27,325) *E. coli* X1776 (ATCC No. 31,337), and *E. coli* B; however many other strains of *E. coli*, such as HB101, JM101, NMS22, NMS38, NMS39, and many other species and genera of prokaryotes including bacilli such as *Bacillus subtilis*, other enterobacteriaceae such as *Salmonella typhimurium* or *Shigella marsecans*, and various *Pseudomonas* species 20 may all be used as hosts. Prokaryotic host cells or other host cells with rigid cell walls are preferably transformed using the calcium chloride method as described in section 1.82 of Sambrook et al., *supra*. Alternatively, electroporation may be used for transformation of these cells. Prokaryote transformation techniques are set forth in Dower, W. J., in *Genetic Engineering, Principles and Methods*, 12:273-296, 25 Plenum Publishing Corp. (1990); Hanahan et al., *Meth. Enzymol.*, 204:63 (1991).

As a representative example, cDNA sequences encoding dirigent protein or pinoresinol/lariciresinol reductase may be transferred to the (His)₆-Tag pET vector commercially available (from Novagen) for overexpression in *E. coli* as heterologous host. This pET expression plasmid has several advantages in high level heterologous expression systems. The desired cDNA insert is ligated in frame to plasmid vector sequences encoding six histidines followed by a highly specific protease recognition site (thrombin) that are joined to the amino terminus codon of the target protein. The histidine "block" of the expressed fusion protein promotes very tight binding to immobilized metal ions and permits rapid purification of the recombinant protein by immobilized metal ion affinity chromatography. The histidine leader sequence is

then cleaved at the specific proteolysis site by treatment of the purified protein with thrombin, and the dirigent protein or pinoresinol/lariciresinol reductase eluted. This overexpression-purification system has high capacity, excellent resolving power and is fast, and the chance of a contaminating *E. coli* protein exhibiting similar binding behavior (before and after thrombin proteolysis) is extremely small.

As will be apparent to those skilled in the art, any plasmid vectors containing replicon and control sequences that are derived from species compatible with the host cell may also be used in the practice of the invention. The vector usually has a replication site, marker genes that provide phenotypic selection in transformed cells, one or more promoters, and a polylinker region containing several restriction sites for insertion of foreign DNA. Plasmids typically used for transformation of *E. coli* include pBR322, pUC18, pUC19, pUC18, pUC19, and Bluescript M13, all of which are described in Sections 1.12-1.20 of Sambrook et al., *supra*. However, many other suitable vectors are available as well. These vectors contain genes coding for ampicillin and/or tetracycline resistance which enables cells transformed with these vectors to grow in the presence of these antibiotics.

The promoters most commonly used in prokaryotic vectors include the β -lactamase (penicillinase) and lactose promoter systems (Chang et al., *Nature* 375:615 (1978); Iakura et al., *Science* 198:1056 (1977); Goeddel et al., *Nature* 281:544 (1979)) and a tryptophan (trp) promoter system (Goeddel et al., *Nucl. Acids Res.* 8:4057 (1980); EPO Appl. Publ. No. 36,776), and the "alkaline" phosphatase systems. While these are the most commonly used, other microbial promoters have been utilized, and details concerning their nucleotide sequences have been published, enabling a "skilled worker" to ligate them "functionally" into plasmid vectors (see, e.g., Beilis et al., *Cell* 20:269 (1980)).

Many eukaryotic proteins normally secreted from the cell contain an endogenous secretion signal sequence as part of the amino acid sequence. Thus, proteins normally found in the cytoplasm can be targeted for secretion by linking a signal sequence to the protein. This is readily accomplished by ligating DNA encoding a signal sequence to the 5' end of the DNA encoding the protein and then expressing this fusion protein in an appropriate host cell. The DNA encoding the signal sequence may be obtained as a restriction fragment from any gene encoding a protein with a signal sequence. Thus, prokaryotic, yeast, and eukaryotic signal sequences may be used herein, depending on the type of host cell utilized to practice the invention. The DNA and amino acid sequence encoding the signal sequence

portion of several eukaryotic genes including, for example, human growth hormone, proinsulin, and proalbulin are known (see Stryer, *Biochemistry* W.H. Freeman and Company, New York, NY, p. 769 (1988)), and can be used as signal sequences in appropriate eukaryotic host cells. Yeast signal sequences, as for example, said phosphatase (Arima et al., *Nucleic Acids Res.* 11:1657 (1983)), alpha-factor, alkaline phosphatase and invertase may be used to direct secretion from yeast host cells.

Prokaryotic signal sequences from genes encoding, for example, LamB or Ompf (Wong et al., *Gene* 68:193 (1988)), MalE, PhoA, or beta-lactamase, as well as other genes, may be used to target proteins from prokaryotic cells into the culture medium.

Trafficking sequences from plants, animals and microbes can be employed in the practice of the invention to direct the gene product to the cytoplasm, endoplasmic reticulum, mitochondria or other cellular components, or to target the protein for export to the medium. These considerations apply to the overexpression of pinoresinol/lariciresinol reductase or dirigent protein, and to direction of expression within cells or intact organisms to permit gene product function in any desired location.

The construction of suitable vectors containing DNA encoding replication sequences, regulatory sequences, phenotypic selection genes and the dirigent protein DNA or pinoresinol/lariciresinol reductase DNA of interest are prepared using standard recombinant DNA procedures. Isolated plasmids and DNA fragments are cleaved, tailored, and ligated together in a specific order to generate the desired vectors, as is well known in the art (see, for example, Sambrook et al., *supra*).

As discussed above, pinoresinol/lariciresinol reductase variants, or dirigent protein variants, are preferably produced by means of mutation(s) that are generated using the method of site-specific mutagenesis. This method requires the synthesis and use of specific oligonucleotides that encode both the sequence of the desired mutation and a sufficient number of adjacent nucleotides to allow the oligonucleotide to stably hybridize to the DNA template.

A dirigent protein gene and/or pinoresinol/lariciresinol reductase gene, or an antisense nucleic acid fragment complementary to all or part of a dirigent protein gene or pinoresinol/lariciresinol reductase gene, may be introduced, as appropriate, into any plant species for a variety of purposes including, but not limited to: altering or improving the color, texture, durability and pest-resistance of wood tissue, especially heartwood tissue; reducing the formation of lignans and/or lignins in plant species, such as corn, which are useful as animal fodder, thereby enhancing the

availability of the cellulose fraction of the plant material to the digestive system of animals ingesting the plant material; reducing the lignan/lignin content of plant species utilized in pulp and paper production, thereby making pulp and paper production easier and cheaper; improving the defensive capability of a plant against predators and pathogens by enhancing the production of defensive lignans or lignins; the alteration of other ecological interactions mediated by lignans or lignins; producing elevated levels of optically-pure lignan enantiomers as medicines or food additives; introducing, enhancing or inhibiting the production of dirigent proteins or pinoresinol/lariciresinol reductases, or the production of pinoresinol or lariciresinol and their derivatives. A dirigent protein and/or pinoresinol/lariciresinol reductase gene may be introduced into any organism for a variety of purposes including, but not limited to: introducing, enhancing or inhibiting the production of dirigent protein and/or pinoresinol/lariciresinol reductase, or the production of pinoresinol or lariciresinol and their derivatives.

The foregoing may be more fully understood in connection with the following representative examples, in which "Plasmids" are designated by a lower case p followed by an alphanumeric designation. The starting plasmids used in this invention are either commercially available, publicly available on an unrestricted basis, or can be constructed from such available plasmids using published procedures. In addition, other equivalent plasmids are known in the art and will be apparent to the ordinary artisan.

"Digestion", "cutting" or "cleaving" of DNA refers to catalytic cleavage of the DNA with an enzyme that acts only at particular locations in the DNA. These enzymes are called restriction endonucleases, and the site along the DNA sequence where each enzyme cleaves is called a restriction site. The restriction enzymes used in this invention are commercially available and are used according to the instructions supplied by the manufacturers. (See also Sections 1.60-1.61 and Sections 3.38-3.39 of Sambrook et al., *supra*.)

"Recovery" or "isolation" of a given fragment of DNA from a restriction digest means separation of the resulting DNA fragment on a polyacrylamide or an agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. This procedure is known generally. For example, see Lawn et al.

(*Nucleic Acids Res.* 9:6103-6114 (1982)), and Goeddel et al. (*Nucleic Acids Res.*, *supra*).

The following examples merely illustrate the best mode now contemplated for practicing the invention, but should not be construed to limit the invention. All literature citations herein are expressly incorporated by reference.

EXAMPLE 1

Purification of Dirigent Protein from *Forsythia intermedia*

Plant Materials. *Forsythia intermedia* plants were either obtained from Bailey's Nursery (var. Lynwood Gold, St. Paul, MN), and maintained in Washington State University greenhouse facilities, or were gifts from the local community.

Initial Extraction and Ammonium Sulfate Precipitation. Solubilization of bound proteins was carried out at 4°C. Frozen *Forsythia intermedia* stems (2 kg) were pulverized in a Waring Blender (Model CB6) in the presence of liquid nitrogen. The resulting powder was homogenized with 0.1 M KH₂PO₄-K₂HPO₄ buffer (pH 7.0, 4 liters) containing 5 mM dithiothreitol, and filtered through four layers of cheesecloth. The insoluble residue was consecutively extracted, with continuous agitation at 250 rpm, as follows: with chilled (-20°C) re-distilled acetone (4 liters, 3 x 30 min); 0.1 M KH₂PO₄-K₂HPO₄ buffer (pH 6.5) containing 0.1% β-mercaptoethanol (solution A, 8 liters, 30 min); solution A containing 1% Triton X100 (8 liters, 4 hours) and finally solution A (8 liters, 16 hours). Between each extraction, the residue was filtered through one layer of Miracloth (Calbiochem). Solubilization of the (+)-pinoresinol forming system was achieved by mechanically stirring the residue in solution A containing 1 M NaCl (8 liters, 4 hours). The homogenate was decanted and the resulting solution consecutively filtered through Miracloth (Calbiochem) and glass fiber (G6; Fisher Sci.). The filtrate was concentrated in an Amicon cell (Model 2000; YM 30 membrane) to a final volume of ~800 ml, and subjected to (NH₄)₂SO₄ fractionation. Proteins precipitating between 40 and 80% saturation were recovered by centrifugation (15,000g, 30 min) and the (NH₄)₂SO₄ pellet stored at -20°C until required.

Mono S Column Chromatography. Purification of 78-kD dirigent protein and

partial purification of oxidase. The ammonium sulfate pellet (obtained from 2 kg of *F intermedia* stems) was reconstituted in 40 mM MES

[2-(N-Morpholino)ethanesulfonic acid] buffer, adjusted to pH 5.0 with 6 M NaOH

(solution B, 30 ml), the slurry being centrifuged (3,600g, 5 min), and the supernatant

dialyzed overnight against solution B (4 liters). The dialyzed extract was filtered

(0.22 μ m) and the sample (35 to 40 mg proteins) was applied to a MonoS HR5/5 (50 mm by 5 mm) column equilibrated in solution B at 4°C. After eluting (flow rate 5 $\text{ml min}^{-1} \text{cm}^{-2}$) with solution B (13 ml), proteins were desorbed with Na_2SO_4 in solution B, using a linear gradient from 0 to 100 mM in 8 ml and holding at this concentration for 32 ml, then implementing a series of step gradients at 133 mM for 50 ml, 166 mM for 50 ml, 200 mM for 40 ml, 233 mM for 40 ml and finally 333 mM Na_2SO_4 for 40 ml. Fractions capable of forming (+)-pinoresinol from *E*-coniferyl alcohol were eluted with 333 mM Na_2SO_4 , combined and stored (-80°C) until needed.

POROS SP-M Matrix Column Chromatography (First Column). Fractions from 15 individual elutions from the MonoS HR5/5 column (33 mM Na_2SO_4) were combined (18.5 mg proteins, 180 ml) and dialyzed overnight against solution C. The dialyzed enzyme solution (190 ml) was filtered (0.22 μ m) and an aliquot (47 ml) was applied to the POROS SP-M column. All separations on a POROS SP-M matrix (100 mm by 4.6 mm), previously equilibrated in 25 mM MES-HEPES-sodium acetate buffer (pH 5.0, solution C), were performed at a flow rate of 60 $\text{ml min}^{-1} \text{cm}^{-2}$ and at room temperature. After elution with solution C (12 ml), the proteins were desorbed with a linear Na_2SO_4 gradient (0 to 0.7 M in 66.5 ml) in solution C, whereupon the concentration established was held for an additional 16.6 ml. Under these conditions, separation of four fractions (I, II, III and IV) was achieved at -40, 47, 55 and 61 mS, respectively. This purification step was repeated three times with the remaining dialyzed enzymatic extract, and fractions I, II, III and IV from each experiment were separately combined. When protease inhibitors (that is, phenylmethanesulfonyl fluoride (0.1 mmol ml^{-1}), EDTA (0.5 nmol ml^{-1}), pepstatin A (1 $\mu\text{g ml}^{-1}$), and antipain (1 $\mu\text{g ml}^{-1}$)) were added during the solubilization and all subsequent purification stages, no differences were observed in the elution profiles of fractions I, II, III, and IV.

POROS SP-M Matrix Column Chromatography (Second Column). Fraction I from the first POROS SP-M Matrix column chromatography step (2.62 mg proteins, 40 ml, -24.6 mS) was diluted in filtered, cold distilled water until the conductivity reached ~8 mS (final volume = 150 ml). The diluted protein solution was then applied onto a POROS SP-M column (100 mm by 4.6 mm). After elution with solution C (12 ml), fraction I was desorbed using a linear Na_2SO_4 gradient from 0 to 0.25 M in 20 ml, whereupon the concentration established was held for another 35 25 ml. This was followed by another linear Na_2SO_4 gradient from 0.25 to 0.7 M in

26 ml which was then held at 0.7 M for an additional 16.6 ml. Fractions eluted at ~30 mS (the ionic strength of the eluent was measured with a flow-through detector) were combined (15 ml, 1.3 mg), diluted with water and rechromatographed. The resulting protein (eluted at -30 mS with the gradient described above) was stored (-80°C) until needed.

Gel filtration. An aliquot from fraction I (595.5 μ g proteins, 3 ml, eluted at ~30 mS), was concentrated to 0.6 ml (Centricon 10, Amicon) and loaded onto a S200 (73.2 cm by 1.6 cm, Pharmacia-LKB) gel chromatographic column equilibrated in 0.1 M MES-HEPES-sodium acetate buffer (pH 5.0) containing 50 mM Na_2SO_4 at 10 4°C. An apparently homogenous 78-kD dirigent protein (242 μ g) was eluted (flow rate 0.25 $\text{ml min}^{-1} \text{cm}^{-2}$) as a single component at 133 ml (V₀ = 105 ml). Molecular weights were estimated by comparison of their elution profiles with the standard protein, β -amylase (200,000), alcohol dehydrogenase (150,000), bovine serum albumin (66,000), ovalbumin (45,000), carbonic anhydrase (29,000) and cytochrome c (12,400).

EXAMPLE 2

Characterization of the Purified Dirigent Protein

Molecular Weight and Isoelectric Point Determination. Polyacrylamide gel electrophoresis (PAGE) was performed in Laemmli's buffer system with gradient (4 to 15% acrylamide, Bio-Rad) gels under denaturing and reducing conditions. Proteins were visualized by silver staining. Gel filtration (S200) chromatography of fraction I gave a protein of native molecular weight ~78 kD, whereas SDS-polyacrylamide gel electrophoresis showed a single band at ~27 kD, suggesting that the native protein exists as a trimer. Isoelectric focusing of the native protein on a 20 25 polyacrylamide gel (pH 3 to 10 gradient) revealed the presence of six bands. After isoelectric focusing, each of these bands was electroblotted onto a polyvinylidene fluoride (PVDF) membrane and subjected to amino terminal sequencing, which established that all had similar sequences indicating a series of isoforms. The ultraviolet-visible spectrum of the protein had only a characteristic protein oxidative center.

Assay of the Ability of the Purified Dirigent Protein to Form (+)-Pinoresinol from *E*-Coniferyl alcohol. The four fractions (I to IV) from the first POROS SP-M

chromatographic step (Example 1) were individually rechromatographed, with each fraction subsequently assayed for (+)-pinoresinol-forming activity with *E*-[9-³H]coniferyl alcohol as substrate for one hour. Fraction I (containing dirigent protein) had very little (+)-pinoresinol-forming activity (<5% of total activity loaded onto the POROS SP M column), whereas fraction III catalyzed nonspecific oxidative coupling to give (\pm)-dehydrononiferol alcohols, (\pm)-pinoresinols, and (\pm)-erythro/threo guaiacylglycerol 8-O-4'-coniferyl alcohol ethers. Thus, Fraction III appeared to contain an endogenous plant oxygenating protein.

Although the putative oxidase preparation (Fraction III) was not purified to electrophoretic homogeneity, the electron paramagnetic resonance (EPR) spectrum of this protein preparation resembled that of a typical plant laccase, *i.e.*, a class of naturally-occurring plant oxygenase proteins. We then studied the fate of *E*-[9-³H]coniferyl alcohol (2 μ mol ml⁻¹, 14.7 kBq) in the presence of, respectively, the oxidase (fraction III), the 78-kD dirigent protein (Fraction I), and both fraction III and the 78-kD protein together. With the fraction III preparation alone, only nonspecific bimolecular radical coupling occurs to give (\pm)-dehydrononiferol alcohols, (\pm)-pinoresinols and (\pm)-erythro/threo guaiacylglycerol 8-O-4'-coniferyl alcohol ethers. With the 78-kD protein by itself, however, a small amount of (+)-pinoresinol formation (<5% over 10 hours) was observed, this being presumed to result from residual traces of oxidizing capacity in the preparation. When both fraction III and the 78-kD protein were combined, full catalytic activity and regio- and stereo-specificity in the product was reestablished, whereby essentially only (+)-pinoresinol was formed. Additionally, with fraction III alone, and when fraction III was combined with the 78-kD protein, the rates of substrate depletion and dimeric product formation were nearly identical. Moreover, essentially no turnover of the dimeric lignan products occurred in either case in the presence of the oxidase, over the time-period (8 hours) examined: subsequent dimer oxidation does not occur when *E*-coniferyl alcohol, the preferred substrate, is still present in the assay mixture.

The 78-kD protein therefore appears to determine the specificity of the bimolecular phenoxy radical coupling reaction.

Gel filtration studies were also carried out with mixtures of the dirigent and fraction III proteins, in order to establish if any detectable protein-protein interaction might account for the stereoselectivity. But no evidence in support of complex formation (*i.e.*, to higher molecular size entities) was observed.

EXAMPLE 3

Effect of the 78-kD Dirigent Protein on

Plant Laccase-Catalyzed Monolignol Coupling

E-coniferyl alcohol coupling assay. *E*-[9-³H]Coniferyl alcohol (4 μ mol ml⁻¹, 29.3 kBq) was incubated with a 120-kD laccase (previously purified from *Forsythia intermedia* stem tissue) over a 24-hour period, in the presence and absence of the dirigent protein, as follows. Each assay consisted of *E*-[9-³H]coniferyl alcohol (4 μ mol ml⁻¹, 29.3 kBq, 7.3 MBq mole liter⁻¹; or 2 μ mol ml⁻¹, 14.7 kBq with fraction III), the 78-kD dirigent protein, an oxidase or oxidant, or both [final concentrations: 770 pmol ml⁻¹ dirigent protein; 10.7 pmol protein ml⁻¹ *Forsythia* laccase; 12 μ g protein ml⁻¹ fraction III; 0.5 μ mol ml⁻¹ FMN; 0.5 μ mol ml⁻¹ FAD; 1 and 10 μ mol ml⁻¹ ammonium peroxodisulfate] in buffer (0.1 M MES-HEPES-sodium acetate, pH 5.0) to a total volume of 250 μ l. The enzymatic reaction was initiated by addition of *E*-[9-³H]coniferyl alcohol. Controls were performed in the presence of buffer alone.

After one hour incubation at 30 °C while shaking, the assay mixture was extracted with ethyl acetate (EtOAc, 500 μ l) containing (\pm)-pinoresinols (7.5 μ g), (\pm)-dehydrononiferol alcohols (3.5 μ g) and erythro/threo (\pm)-guaiacylglycerol 8-O-4'-coniferyl alcohol ethers (7.5 μ g) as radiochemical carriers and ferulic acid (15.0 μ g) as an internal standard. After centrifugation (13,800g, 5 min), the EtOAc soluble components were removed and the extraction procedure repeated with EtOAc (500 μ l). The EtOAc soluble components from each assay were combined, the solutions evaporated to dryness *in vacuo*, redissolved in methanol/water solution (1:1; 100 μ l) with an aliquot (50 μ l) thereof subjected to reversed-phase column chromatography (Waters, Nova-Pak C₁₈, 150 mm by 3.8 mm). The elution conditions were as follows: acetonitrile/3% acetic acid in H₂O (5:95) from 0 to 5 min, then linear gradients to ratios of 10:90 between 5 and 20 min, then to 20:80 between 20 and 45 min and finally to 50:50 between 45 and 60 min, at a flow rate of 8.8 ml min⁻¹ cm⁻².

Fractions corresponding to *E*-coniferyl alcohol, erythro/threo (\pm)-guaiacylglycerol 8-O-4'-coniferyl alcohol ethers, (\pm)-dehydrononiferol alcohols and (\pm)-pinoresinols were individually collected, aliquots removed for liquid scintillation counting, and the remainder freeze-dried. Pinoresinol-*c*-containing fractions were redissolved in methanol (100 μ l) and subjected to chiral column chromatography (Daicel, Chiracel OD, 50 mm by 4.6 mm) with a solution of

hexanes and ethanol (1:1) as the mobile phase (flow rate 3 ml min⁻¹ cm⁻²), whereas dehydrodiconiferyl alcohol fractions were subjected to Chiracel OD (250 mm by 4.6 mm) column chromatography eluted with a solution of hexanes and isopropanol (9:1) as the mobile phase (flow rate 2.4 ml min⁻¹ cm⁻²), the radioactivity of the eluent being measured with a flow-through detector (Radiomatic, Model A120).

5 *Results of E-coniferyl alcohol coupling assay.* Incubation with laccase alone gave only racemic dimeric products with (±)-dehydrodiconiferyl alcohols predominating. In the presence of the dirigent protein, however, the process was now primarily stereoselective, affording (+)-pinoresinol, rather than being nonspecific as 10 observed when only laccase was present. The rates of both E-coniferyl alcohol (substrate) depletion and the formation of the dimeric lignans were similar with and without the dirigent protein. A substantial difference was noted in the subsequent turnover of the lignan products observed after E-coniferyl alcohol depletion. With the laccase alone no turnover occurred, but when both proteins were present the 15 disappearance of the products was significant. In order to understand the difference, assays were conducted where bovine serum albumin (BSA) and ovalbumin were individually added to the laccase-containing solutions at levels matching the weight concentrations of the dirigent protein. In this way, it was established that the 20 differences in product turnover were simply due to stabilization of laccase activity at the higher protein concentrations, although interestingly the dirigent protein, BSA and ovalbumin afforded somewhat different degrees of protection. The findings were quite comparable when a fungal laccase (from *Trametes versicolor*) was used in place of the plant laccase. When the oxidizing capacity (i.e., laccase concentration) was lowered five-fold, only (+)-pinoresinol formation was observed. Thus, complete 25 stereoselectivity is preserved when the oxidative capacity does not exceed a point where the dirigent protein is saturated.

5 *Stereoselective E-coniferyl alcohol coupling.* Assays were also conducted with E-[9-2H₂, OC₂H₃]coniferyl alcohol and the dirigent protein in the presence of laccase as follows. E-[9-2H₂, OC₂H₃]coniferyl alcohol (2 μmol ml⁻¹) was incubated 10 in the presence of dirigent protein (770 pmol ml⁻¹), the purified plant laccase (4.1 pmol ml⁻¹) and buffer (0.1 M MES-HEPES-sodium acetate, pH 5.0) in a total 30 volume of 250 μl. After one hour incubation, the reaction mixture was extracted with EtOAc, but with the addition of an internal standard and radiochemical carriers omitted. After reversed-phase column chromatography, the enzymatically formed 35 pinoresinol was collected, freeze-dried, redissolved in methanol (100 μl) and

subjected to chiral column chromatography (Daicel, Chiracel OD, 50 mm by 4.6 mm) with detection at 280 nm and analysis by mass spectral fragmentation in the EI mode (Waters, Integrity System). Liquid chromatography-mass spectrometry (LC-MS) analysis of the resulting (+)-pinoresinol (>99% enantiomeric excess) gave a molecular ion with a mass to charge ratio (m/z) 368, thus establishing the presence of 10-2H atoms and verifying that together the laccase and dirigent protein catalyzed stereoselective coupling of E-[9-2H₂, OC₂H₃]coniferyl alcohol.

Other auxiliary one-electron oxidants can also facilitate stereoselective coupling with the dirigent protein. Ammonium peroxydisulfate readily undergoes homolytic cleavage (A. Usaitis, R. Makuska, *Polymer* 35:4896 (1994)) and is routinely used as an one-electron oxidant in acrylamide polymerization. Ammonium peroxydisulfate was first incubated with E-[9-3H]coniferyl alcohol (4 μmol ml⁻¹, 29.3 kBq) for 6 hours using the E-coniferyl alcohol coupling assay procedure described above. Nonspecific bimolecular radical coupling was observed, to afford 15 predominantly (±)-dehydrodiconiferyl alcohols as well as the other racemic lignans (Table 1). However, when the dirigent protein was added, the stereoselectivity of coupling was dramatically altered to give primarily (+)-pinoresinol at both concentrations of oxidant, together with small amounts of racemic lignans. This 20 established that even an inorganic oxidant, such as ammonium peroxydisulfate, could promote (+)-pinoresinol synthesis in the presence of the dirigent protein, even if it was not oxidatively as selective toward the monolignol as was the fraction III oxidase or laccase.

Oxidant (μ mol ml $^{-1}$)	Diligent protein in dimer equivalents	E-Confiferyl alcohol		(-)-Quinic-glyceral		(-)-Dehydro-		Peroxidisulfate		Ammonium peroxidisulfate		Ammonium peroxidisulfate		Difigent (μ mol ml $^{-1}$)	
		(μ mol ml $^{-1}$)	(μ mol ml $^{-1}$)	(μ mol ml $^{-1}$)	(μ mol ml $^{-1}$)	(μ mol ml $^{-1}$)	(μ mol ml $^{-1}$)								
Ammonium peroxidisulfate (μ mol ml $^{-1}$)	absent	250 \pm 55	6 \pm 0	13 \pm 1	0	130 \pm 10	149 \pm 11								
Ammonium peroxidisulfate (μ mol ml $^{-1}$)	present	250 \pm 30	90 \pm 4	250 \pm 10	135 \pm 4	0	475 \pm 17								
Ammonium peroxidisulfate (μ mol ml $^{-1}$)	absent	860 \pm 30	6 \pm 0	13 \pm 1	0	130 \pm 10	149 \pm 11								
Ammonium peroxidisulfate (μ mol ml $^{-1}$)	present	1030 \pm 25	30 \pm 1	90 \pm 3	0	450 \pm 10	570 \pm 14								
Difigent (μ mol ml $^{-1}$)	absent	1030 \pm 25	30 \pm 1	90 \pm 3	0	450 \pm 10	570 \pm 14								
Diligent protein in dimer equivalents	absent	200 \pm 4	10 \pm 1	35 \pm 2	16 \pm 0	0	61 \pm 3								
Diligent protein in dimer equivalents	present	1030 \pm 25	30 \pm 1	90 \pm 3	0	450 \pm 10	570 \pm 14								

Table 1.
Effect of diligent protein on product distribution from E-Confiferyl alcohol
oxidized by ammonium peroxidisulfate (6 hour assay).

Effect of Other Oxygenating Agents on the Stereospecific Conversion of E-Confiferyl Alcohol to (+)-Pinoresinol. The effects of incubating E-Confiferyl alcohol (4 μ mol ml $^{-1}$, 29.3 kBq) with flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) were investigated since, in addition to their roles as enzyme cofactors, they can also oxidize various organic substrates (T.C. Bruce, *Acc. Chem. Res.* 13:256 (1980)). E-[9- 3 H]Confiferyl alcohol was respectively incubated with FMN and FAD for 48 hours. To obtain the FMN, snake (*Naja naja atra*, Formosan cobra) venom was added to a solution of FAD (5 μ mol ml $^{-1}$ in H₂O) and, after 20 min incubation at 30°C, the enzymatically formed FMN was separated from the protein mixture by filtration through a Centriflo 10 (Amicon) microconcentrator. In every instance, E-Confiferyl alcohol oxidation was more rapid in the presence of FMN than FAD. Although these differences between the FMN and FAD catalyzed rates of E-Confiferyl alcohol oxidation were not anticipated, a consistent pattern was sustained: racemic lignan products were obtained, with the (\pm -)dehydrodiconiferyl alcohols predominating as before. When the time courses were repeated in the presence of the dirigent protein, a dramatic change in stereoselectivity was observed, where essentially only (+)-pinoresinol formation occurred. Again, the rates of E-Confiferyl alcohol depletion, when adjusted for the traces of residual oxidizing capacity (<5% over 10 hours) in the dirigent protein preparation, were dependent only upon [FMN] and [FAD], as were the total amounts of dimers formed. When full depletion of E-Confiferyl alcohol occurs, the corresponding lignan dimers can begin to undergo oxidative changes as a function of time; specifically, FMN is able subsequently to oxidize pinoresinol, in open solution, after the E-Confiferyl alcohol has been fully depleted.

Investigation of Substrate-Specific Stereoselectivity. The coupling stereoselectivity was substrate specific. Neither *E-p-[9- 3 H]coumaryl* (4 μ mol ml $^{-1}$, 44.5 kBq) or *E-[8- 14 C]sinapyl* alcohols (4 μ mol ml $^{-1}$, 8.3 kBq), which differ from E-Confiferyl alcohol only by a methoxyl group substituent on the aromatic ring, yielded stereoselective products when incubated for 6 hours with FMN and ammonium peroxidisulfate respectively, in the presence and absence of the dirigent protein. Incubations were carried out as described above with the following modifications: *E-p-[9- 3 H]coumaryl* (4 μ mol ml $^{-1}$, 44.5 kBq) or *E-[8- 14 C]sinapyl* alcohols (4 μ mol ml $^{-1}$, 8.3 kBq) were used as substrates and, after 6 hour incubation at 30°C, the reaction mixture was extracted with EtOAc but without addition of radiochemical carriers. E-Sinapyl alcohol readily underwent coupling to afford

syringaresinol, but chiral HPLC analysis revealed that the resulting products were, in every instance, racemic (Table 2). Interestingly, by itself, the 78-kD dirigent protein preparation catalyzed a low level of dimer formation, as previously noted, but only gave rise to racemic (±)-syringaresinol formation, which is presumably a consequence of the residual traces of contaminating oxidizing capacity present in the protein preparation.

In an analogous manner, no stereoselective coupling was observed with *E*-pincony alcohol as substrate. That is, only *E*-coniferyl alcohol undergoes stereoselective coupling in the presence of the dirigent protein. Given the marked substrate specificity of the dirigent protein for *E*-coniferyl alcohol, it will be of considerable interest to determine, in the future, how it differs from that affording (±)-syringaresinol in *Eucalyptus illicoides* (T. Deyama, *Chem. Pharm. Bull.* 31, 2993 (1983)).

Table 2.
Effect of Dirigent Protein on coupling of *E*-syringaresinol (6 hour assay).

FMN	Dirigent protein (770 pmol ml ⁻¹)	E-Syringaresinol dimer equivalents (nmol ml ⁻¹)	Racemic (±)-syringaresinols (nmol ml ⁻¹)	Syringaresinol (±)-syringaresinols (nmol ml ⁻¹)	
				absent	present
(0.5 μmol ml ⁻¹)	absent	570 ± 100	1290 ± 40		
	present	610 ± 110	340 ± 40		
Ammonium peroxydisulfate (10 μmol ml ⁻¹)	absent	1400 ± 120	1020 ± 40		
	present			1520 ± 10	1060 ± 30
Dirigent protein	absent			110 ± 10	50 ± 10

Although the inventors do not intend to be bound by any particular mechanism for stereoselective coupling, three distinct possibilities can be envisaged. The most likely is that the oxidase or oxidant generates free-radical species from *E*-coniferyl alcohol, and that the latter are the true substrates that bind to the dirigent protein prior to coupling. The other two possibilities would require that *E*-coniferyl alcohol molecules are bound and oriented on the dirigent protein, thereby ensuring that only (+)-pinoresinol formation occurs upon subsequent oxidative coupling; this

could occur either if both substrate phenolic hydroxyl groups were exposed so that they could readily be oxidized by an oxidase or oxidant, or if an electron transfer mechanism were operative between the oxidase or oxidant and an electron acceptor site or sites on the dirigent protein.

Among the three alternative mechanisms, three lines of evidence suggest "capture" of phenoxyl radical intermediates by the dirigent protein. First, the rates of both substrate depletion and product formation are largely unaffected by the presence of the dirigent protein. If capture of the free-radical intermediates is the operative mechanism, the dirigent protein would only affect the specificity of coupling when single-electron oxidation of coniferyl alcohol is rate-determining. Second, an electron transfer mechanism is currently ruled out, since we observed no new ultraviolet-visible chromophores in either the presence or absence of an auxiliary oxidase or oxidant, under oxidizing conditions. Third, preliminary kinetic data (as disclosed in Example 4) support the concept of free-radical capture based on the formal values of Michaelis constant (K_m) and maximum velocity (V_{max}) characterizing the conversion of *E*-coniferyl alcohol into (+)-pinoresinol, with the dirigent protein alone and in the presence of the various oxidases or oxidants.

EXAMPLE 4

Kinetic Characterization of the Conversion of *E*-Coniferyl Alcohol to (+)-pinoresinol in the Presence of Dirigent Protein and an Oxygenating Agent.

Assays were carried out as described in Example 3 by incubating a series of *E*-[³H]coniferyl alcohol concentrations (between 8.00 and 0.13 μmol ml⁻¹, 7.3 MBq mole liter⁻¹) with dirigent protein (770 pmol ml⁻¹) alone and in presence of *Forsythia* laccase (2.1 pmol ml⁻¹), fraction III (12 μg protein ml⁻¹), or FMN (0.5 μmol ml⁻¹). Assays with dirigent protein, in presence or absence of FMN, were incubated at 30°C for 1 hour, whereas assays with *Forsythia* laccase or fraction III in presence or absence of dirigent protein were incubated at 30°C for 15 min. If free-radical capture by the dirigent protein is the operative mechanism, the Michaelis-Menten parameters obtained will only represent formal rather than true values, because the highest free-energy intermediate state during the conversion of *E*-coniferyl alcohol into (+)-pinoresinol is still unknown and the relation between the concentration of substrate and that of the corresponding intermediate free-radical in open solution has not been delineated.

Bearing these qualifications in mind, we estimated formal K_m and V_{max} values for the dirigent protein preparation. As noted earlier, it was capable of

35

engendering formation of low levels of both (+)-pinoresinol from *E*-coniferyl alcohol, and racemic (±)-syringaresinols from *E*-sinapyl alcohol, because of traces of contaminating oxidizing capacity. With this preparation (Table 3), a formal K_m of 10 ± 6 mM and V_{max} of 0.02 ± 0.02 mol s⁻¹ mol⁻¹ were obtained. However, with addition of fraction III, laccase, and FMN, the formal K_m values (mM) were reduced to 1.6 ± 0.3, 0.100 ± 0.003, and 0.10 ± 0.01, respectively, whereas the V_{max} values were far less affected at these concentrations of auxiliary oxidase/oxidant.

Formal K_m and V_{max} values were calculated for the laccase and fraction III oxidase with respect to *E*-coniferyl alcohol conversion into the three racemic lignans. However, no direct comparisons can be made to the 78-kD protein, since the formal K_m values involve only the corresponding oxidases. For completeness, the K_m (mM) and V_{max} (mol s⁻¹ mol⁻¹ enzyme) were as follows: with respect to the laccase, 0.200 ± 0.001 and 3.9 ± 0.2 for (±)-erythro/threo guaiacylglycerol 8-O-4'-coniferyl alcohol ethers, 0.3000 ± 0.0003 and 13.1 ± 0.6 for (±)-dehydrononiferol alcohols, 15 and 0.300 ± 0.002 and 7.54 ± 0.50 for (±)-pinoresinols; with respect to the fraction III oxidase (estimated to have a native molecular weight of 80 kDa), 2.2 ± 0.3 and 0.20 ± 0.03 for (±)-erythro/threo guaiacylglycerol 8-O-4'-coniferyl alcohol ethers, 2.2 ± 0.2 and 0.7 ± 0.1 for (±)-dehydrononiferol alcohols, and 3.7 ± 0.7 and 0.6 ± 0.1 for (±)-pinoresinols.

These preliminary kinetic parameters are in harmony with the finding that dirigent protein does not substantially affect the rate of *E*-coniferyl alcohol depletion in the presence of fraction III, laccase and FMN. Both sets of results are together in accord with the working hypothesis that the dirigent protein functions by capturing free-radical intermediates which then undergo stereoselective coupling.

Table 3.
Effect of various oxidants on formal K_m and V_{max} values for the dirigent protein (770 pmol ml⁻¹) during (+)-pinoresinol formation from *E*-coniferyl alcohol.

Oxidase/Oxidant	Formal K_m (mM)	V_{max} (mol s ⁻¹ mol ⁻¹ dirigent protein)
Dirigent protein	10 ± 6	0.02 ± 0.02
Fraction III (12 µg protein ml ⁻¹)	1.6 ± 0.3	0.10 ± 0.03
Laccase (2.07 pmol ml ⁻¹)	0.100 ± 0.003	0.0600 ± 0.0002
FMN (0.5 µmol ml ⁻¹)	0.10 ± 0.01	0.024 ± 0.001

EXAMPLE 5
Cloning of the Dirigent Protein cDNA From *Forsythia intermedia*

Plant Materials - *Forsythia intermedia* plants were either obtained from Bailey's Nursery (var. Lynwood Gold, St. Paul, MN), and maintained in Washington State University greenhouse facilities, or were gifts from the local community.

Materials - All solvents and chemicals used were reagent or HPLC grade. *Taq* thermostable DNA polymerase was obtained from Promega, whereas restriction enzymes were from Gibco BRL (*Hae*III), Boehringer Mannheim (*Sau*3A) and Promega (*Taq*). pT7Blue T-vector and competent NovaBlue cells were purchased from Novagen and radiolabeled nucleotide ([α -32P]dCTP) was from DuPont NEN.

Oligonucleotide primers for polymerase chain reaction (PCR) and sequencing were synthesized by Gibco BRL Life Technologies. GENECLEAN II[®] kits (BIO 101 Inc.) were used for purification of PCR fragments, with the gel-purified DNA concentrations determined by comparison to a low DNA mass ladder (Gibco BRL) in 1.5% agarose gels.

Instrumentation - UV (including RNA and DNA determinations at OD₂₆₀) spectra were recorded on a Lambda 6 UV/VIS spectrophotometer. A Temptronic II thermocycler (Thermolyne) was used for all PCR amplifications. Purification of DNA for sequencing employed a QIAwell Plus plasmid purification system (QIAGEN) followed by PEG precipitation (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1994) *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), with DNA sequences determined using an Applied Biosystems Model 373A automated sequencer. Amino acid sequences were obtained using an Applied Biosystems protein sequencer with on-line HPLC detection, according to the manufacturer's instructions.

Dirigent Protein Amino Acid Sequencing - The dirigent protein N-terminal amino acid sequence (SEQ ID No.:1) was obtained from the purified protein using an Applied Biosystems protein sequencer with on-line HPLC detection. For trypsin digestion, the purified enzyme (150 pmol) was suspended in 0.1 M Tris-HCl (50 µl, pH 8.5, Boehringer Mannheim, sequencing grade), with urea added to give a final concentration of 8 M in 77.5 µl. The mixture was incubated for 15 min at 50°C, following which 100 mM iodoacetamide (2.5 µl) was added, with the whole kept at room temperature for 15 min. Trypsin (1 µg in 20 µl) was then added, with the mixture digested for 24 h at 37°C, following which TFA (4 µl) was added to stop the enzymatic reaction. The resulting mixture was subjected to reversed phase HPLC

analysis (C-8 column, Applied Biosystems), this being eluted with a linear gradient over 2 h from 0 to 100% acetonitrile (in 0.1% TFA) at a flow rate of 0.2 ml/min with detection at 280 nm. Fractions containing individual oligopeptide peaks were collected manually and directly submitted to amino acid sequencing (SEQ ID Nos: 2-7).

Forsythia intermedia stem cDNA Library Synthesis - Total RNA (~300 µg/g fresh weight) was obtained (Dong, Z.D. and Danstan, D.L. (1996) *Plant Cell Reports* 15:516-521) from young green stems of greenhouse-grown *Forsythia intermedia* plants (var. Lynwood Gold). A *Forsythia intermedia* stem cDNA library was constructed using 5 µg of purified poly A+ mRNA (Oligotex-dT™ Suspension, QIAGEN) with the ZAP-*o*-DNA® synthesis kit, the Uni-ZAP™ XR vector and the Gigapack® II Gold packaging extract (Stratagene), with a titer of 1.2 x 10⁶ PFU for the primary library. A portion (30 ml) of the amplified library (1.2 x 10⁶ PFU/ml; 158 ml total) (Sambrook, J. et al., *supra*) was used to obtain pure cDNA library DNA.

(Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1991) *Current Protocols in Molecular Biology*, 2 volumes, Greene Publishing Associates and Wiley-Interscience, John Wiley & Sons, NY) for PCR.

Direct™ Protein DNA Probe Synthesis - The N-terminal and internal peptide amino acid sequences were used to construct the degenerate oligonucleotide primers.

Purified *F. intermedia* cDNA library DNA (5 ng) was used as the template in 100 µl PCR reactions (10 mM Tris-HCl, 50 mM KCl, 0.1% Triton X-100, 2.5 mM MgCl₂, 0.2 mM each dNTP and 2.5 units *Taq* DNA polymerase) with primer PSINT1 (SEQ ID No:8) (100 pmol) and either primer PS17R (SEQ ID No:11) (20 pmol), primer PS12R (SEQ ID No:10) (20 pmol) or primer PS11R (SEQ ID No:9) (20 pmol). PCR amplification was carried out in a thermocycler as follows: 35 cycles of 1 min at 94°C, 2 min at 50°C and 3 min at 72°C with 5 min at 72°C and an indefinite hold at 4°C after the final cycle. Single-primer, template-only and primer-only reactions were performed as controls. PCR products were resolved in 1.5% agarose gels, where a single band (~370, ~155- or ~125-bp, respectively) was observed for each reaction.

To determine the nucleotide sequence of the amplified bands, five 100 µl

PCR reactions were performed as above with PSINT1 (SEQ ID No:8) +PS17R (SEQ ID No:11), PSINT1 (SEQ ID No:8) +PS12R (SEQ ID No:10) and PSINT1 (SEQ ID No:8) +PS11R (SEQ ID No:9) primer pairs. The 5 reactions from each primer pair were concentrated (Microcon™30, Amicon Inc.) and washed with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 2 x 200 µl), with the PCR products subsequently recovered in TE buffer (2 x 50 µl). These were resolved in preparative 1.5% agarose gels. Each gel-purified PCR product (~0.2 pmol) was then ligated into the pTBlue T-vector and transformed into competent Novabio cells, according to the Novagen's instructions. Insert sizes were determined using the rapid boiling lysis and PCR "technique" (with R20mer and U19mer primers) according to the manufacturer's instructions. Restriction analyses were performed to determine whether all inserts from the reactions utilizing each of the foregoing primer pairs were the same; as follows: to 20 µl each of a 100 µl PCR reaction (insert of interest amplified with R20mer (SEQ ID No:74) and U19mer (SEQ ID No:75) primers) were added 4 units *Hae*II, 1.5 units *Sma*II or 5 units *Xba*I restriction enzyme. Restriction digestions were allowed to proceed for 60 min at 37°C for *Hae*II and *Sma*II and at 65°C for *Xba*I reactions. Restriction products were resolved in 1.5% agarose gels giving one restriction group for each insert tested. Five recombinant plasmids from PSINT1 (SEQ ID No:8) +PS17R (SEQ ID No:11) (called pT7PS11-pT7PS15) and 2 recombinant plasmids from PSINT1 (SEQ ID No:8) +PS12R (SEQ ID No:10) (called pT7PS16 and pT7PS17). PCR products were selected for DNA sequencing; all contained the same open reading frame (ORF) (SEQ ID No:69). The divergent protein probe was next constructed as follows: five 100 µl PCR reactions were performed as above with 10 ng pT7PS11 DNA (SEQ ID No:69) with primers PSINT1 (SEQ ID No:8) and PS17R (SEQ ID No:11). Gel-purified pT7PS11 insert (50 ng) was used with Pharmacia's T/QuickPrime® kit and [³²P]dCTP, according to kit instructions, to produce a radiolabeled probe (in 0.1 ml), which was purified over BioSpin 6 columns (Bio-Rad) and added to carrier DNA (0.5 mg/ml sheared salmon sperm DNA (Sigma), 0.9 ml).

Library Screening - 600,000 PFU of *F. intermedia* amplified cDNA library were plated for primary screening, according to Stratagene's instructions. Plaques were blotted onto Magna Nylon membrane circles (Micron Separations Inc.), which were then allowed to air dry. The membranes were placed between two layers of Whatman® 3MM Chr paper. cDNA library phage DNA was fixed to the membranes and denatured in one step by autoclaving for 2 min at 100°C with fast exhaust. The membranes were washed for 30 min at 37°C in 6X standard saline citrate (SSC) and 0.1% SDS and prehybridized for 5 h with gentle shaking at 57-58°C in preheated 6X SSC, 0.5% SDS and 5X Denhardt's reagent (hybridization solution, 300 ml) in a crystallization dish (190 x 75 mm). The [³²P]radiolabeled probe was denatured

35

30 Whatman® 3MM Chr paper. cDNA library phage DNA was fixed to the membranes and denatured in one step by autoclaving for 2 min at 100°C with fast exhaust. The membranes were washed for 30 min at 37°C in 6X standard saline citrate (SSC) and 0.1% SDS and prehybridized for 5 h with gentle shaking at 57-58°C in preheated 6X SSC, 0.5% SDS and 5X Denhardt's reagent (hybridization solution, 300 ml) in a crystallization dish (190 x 75 mm). The [³²P]radiolabeled probe was denatured

-43-

(boiling, 10 min), quickly cooled (ice, 15 min) and added to a preheated fresh hybridization solution (60 ml, 58°C) in a crystallization dish (150 x 75 mm). The prehybridized membranes were next added to this dish, which was then covered with plastic wrap. Hybridization was performed for 18 h at 57-58°C with gentle shaking.

5 The membranes were washed in 4X SSC and 0.5% SDS for 5 min at room temperature, transferred to 2X SSC and 0.5% SDS (at room temperature) and incubated at 57-58°C for 20 min with gentle shaking, wrapped with plastic wrap to prevent drying and finally exposed to Kodak X-OMAT AR film for 24 h at -80°C with intensifying screens. Twenty positive plaques were purified through two more 10 rounds of screening with hybridization conditions as above.

In vivo Excision and Sequencing of Dirigent Protein cDNA-containing Phagemids - Purified cDNA clones were rescued from the phage following Stratagene's *in vivo* excision protocol. Both strands of several different cDNAs that coded for dirigent protein were completely sequenced using overlapping sequencing 15 primers. Two distinct cDNAs were identified, called pPSD_F1 (SEQ ID No:12) and pPSD_F12 (SEQ ID No:14).

Sequence Analysis - DNA and amino acid sequence analyses were performed using the Unix-based GCG Wisconsin Package (Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, 20 Madison, Wisconsin, USA 53711; Rice, P. (1996) *Program Manual for the EGCG Package*, Peter Rice, The Sanger Centre, Hinxton Hall, Cambridge, CB10 1RQ, England) and the ExPASy World Wide Web molecular biology server (Geneva University Hospital and University of Geneva, Geneva, Switzerland).

EXAMPLE 6

Expression of Functional Dirigent Protein in *Spodoptera frugiperda*

Attempts to express functional dirigent protein in *Escherichia coli* failed. Consequently, we expressed the dirigent protein in *Spodoptera frugiperda* utilizing a baculovirus expression system. The full-length 1.2 kb cDNA clone for the dirigent protein (PSD) in *F. intermedia*, containing both the 5' and 3' untranslated regions, 25 was excised from the pBlueScript (Stratagene) derived plasmid pPSD_F11 (SEQ ID No:12) using the restriction endonucleases *Bam*H 1 and *Xba* 1. This 1.2 kb fragment was directionally subcloned into these same restriction sites in the multiple cloning site of the baculovirus transfer vector pBlueBac4 (Invitrogen, San Diego, CA). This produced the 6.0 kb construct pBB4/PSD which generates a non-fusion dirigent protein with translation being initiated at the dirigent protein cDNA 30

start codon. This construct was then co-transfected with linearized Bac-N-Blue DNA (Invitrogen) into *Spodoptera frugiperda* Sf9 cells by the technique of cationic liposome mediated transfection to produce, by means of homologous recombination, the recombinant *Autographa californica* nuclear polyhedrosis viral (AcMNPV) DNA 5 Bac-N-Blue dirigent protein (BBPSD) which was purified from plaques according to procedures described by Invitrogen. The final recombinant AcMNPV-BBPSD contains the PSD gene under the polyhedrin promoter control and the essential sequence needed for replication of the recombinant virus. To verify that the dirigent protein was successfully expressed in the insect cell culture, log phase Sf9 cells infected with the AcMNPV-PSD recombinant viral high titer stock were used to obtain heterologous protein production. Maximal dirigent protein yield occurred by 10 48-70 hours post-infection. As determined by SDS-PAGE and (+)-pinoresinol forming activity, the protein was found secreted into the medium and showed a molecular mass and activity which corresponded to the indigenous protein originally isolated from *Forsythia intermedia*.

EXAMPLE 7

*Isolation of Dirigent Protein Clones from *Thuja plicata* and *Tsuga heterophylla**

The coding region of a *Forsythia* dirigent protein cDNA, pfd-Fil (SEQ ID No:12), was used to screen cDNA libraries from *Thuja plicata* and *Tsuga heterophylla*. The conditions and methods were as disclosed in Example 5, except that hybridization was carried out at 45-50°C. Two dirigent protein cDNAs were isolated from *Tsuga heterophylla* (SEQ ID Nos:16, 18), and eight dirigent protein cDNAs were isolated from *Thuja plicata* (SEQ ID Nos:20, 22, 24, 26, 28, 30, 32, 34).

EXAMPLE 8

*Purification of Pinoresinol/Farinresinol Reductases from *Forsythia Intermedia**

Plant Materials. *Forsythia intermedia* plants were either obtained from Bailey's Nursery (var. Lynwood Gold, St. Paul, MN), and maintained in Washington State University greenhouse facilities, or were gifts from the local community.

20 *Materials.* All solvents and chemicals used were reagent or HPLC grade. Unlabeled (+)-pinoresinols and (±)-farinresinols were synthesized as described (Katayama, T. et al., *Phytochemistry* 33:581-591 (1993)). [4R-3H]NADPH was obtained as previously reported (Chu, A. et al., *J. Biol. Chem.* 268:2026-27033 (1993)) by modification of the procedure of Moran et al. (Moran, R.G. et al., *Anal. Biochem.* 138:196-204 (1984)), and [4R-2H]NADPH was prepared according to

Anderson and Lin (Anderson, J.A., and Lin B.K., *Physichemistry* 32:811-812 (1993)). Yeast glucose-6-phosphate dehydrogenase (Type DX-22; 32. mmol h⁻¹ mg⁻¹) and yeast hexokinase (Type F300, 15.12 mmol⁻¹ mg⁻¹) were purchased from Sigma and dihydrofolate reductase (*Lactobacillus casei*, 33.48 mmol h⁻¹ mg⁻¹) was obtained from Biopure Co. Affi-Gel Blue Gel (100-200 mesh) and Bio-Gel HT Hydroxypatite were purchased from Bio-Rad, whereas Phenyl Sepharose CL-4B, MonoQ HR 5/5, MonoP HR 5/20, Superose 6, Superose 12, Superdex 75, PD-10 columns, molecular weight standards and Polybuffer 74 were obtained from Pharmacia LKB Biotechnology, Inc. Adenosine 2',5'-diphosphate Sepharose and Reactive Yellow 3 Agarose were from Sigma Chemical Co.

Instrumentation. ¹H Nuclear magnetic resonance spectra (300 and 500 MHz) were recorded on Bruker AMX300 and Varian VXR500S spectrometers, respectively, using CDCl₃ as solvent with chemical shifts (δ ppm) reported downfield from tetramethylsilane (internal standard). UV (including RNA and DNA determinations at OD₂₆₀) and mass spectra were obtained on Lambda 6 UV/VIS and VG 7070E (ionizing voltage 70 eV) spectrophotometers, respectively. High performance liquid chromatography was carried out using either reversed-phase (Waters, Novapak C18-250 x 9.3 mm inner diameter) or chiral (Daicel Chiraleel OD or Chiraleel OC, 240 x 4.6 mm inner diameter) columns, with detection at 280 nm (Chu, A., et al., *J. Biol. Chem.* 268:27026-27033 (1993)). Radioactive samples were analyzed in E-coulee (ICN) and measured using a liquid scintillation counter (Packard, Tricarb 2000 CA). Amino acid sequences were obtained using an Applied Biosystems protein sequencer with on-line HPLC detection, according to the manufacturer's instructions.

Enzyme assays. Pinoresinol and lariciresinol reductase activities were assayed by monitoring the formation of (+)-[³H]pinoresinol and (-)-[³H]secoisolariciresinol (Chu, A., et al., *J. Biol. Chem.* 268:27026-27033 (1993)). Briefly, each assay for pinoresinol reductase activity consisted of (\pm)-pinoresinols (5 mM in MeOH, 20 μl), the enzyme preparation at the corresponding stage of purity (100 μl), and buffer (20 mM Tris-HCl, pH 8.0, 110 μl). The enzymatic reaction was initiated by addition of [4R-³H]NADPH (10 mM, 6.79 kBq/mmol in 20 μl of double-distilled H₂O). After 30 min incubation at 30°C with shaking, the assay mixture was extracted with EtOAc (500 μl) containing (\pm)-lariciresinols (20 μg) and (\pm)-secoisolariciresinols (20 μg) as radiochemical carriers. After centrifugation (13,800 x g, 5 min), the EtOAc solubles were removed

and the extraction procedure was repeated. For each assay, the EtOAc solubles were combined with an aliquot (100 μl) removed for determination of its radioactivity using liquid scintillation counting. The remainder of the combined EtOAc solubles was evaporated to dryness in vacuo, reconstituted in MeOH/3% acetic acid in H₂O (30:70, 100 μl) and subjected to reversed phase and chiral column HPLC. Controls were performed using either denatured enzyme (boiled for 10 min) or in the absence of (\pm)-pinoresinols as substrate.

Lariciresinol reductase activity was assayed by monitoring the formation of (-)-[³H]secoisolariciresinol. These assays were carried out exactly as described above, except that (\pm)-lariciresinols (5 mM in MeOH, 20 μl) were used as substrates, with (\pm)-secoisolariciresinols (20 μg) added as radiochemical carriers.

General Procedures for Enzyme Purification. Protein purification procedures were carried out at 4°C with chromatographic eluents monitored at 280 nm, unless otherwise stated. Protein concentrations were determined by the method of Bradford (Bradford, M.M., *Anal. Biochem.* 72:248-254 (1976)) using γ -globulin as standard. Polyacrylamide gel electrophoresis used gradient (4-15%, Bio-Rad) gels under denaturing and reducing conditions, these being performed in Laemmli's buffer system (Laemmli, U.K., *Nature* 227:680-685 (1970)). Proteins were visualized by silver staining (Morrissey, J.H., *Anal. Biochem.* 117:307-310 (1981)).

Preparation of crude extracts. *F. intermedia* stems (20 kg) were harvested, cut into 3-6 cm sections, and stored at -20°C until needed. Batches of stems (2 kg) were frozen in liquid nitrogen and pulverized in a Waring Blender. The resulting powder was homogenized with potassium phosphate buffer (0.1 mM, pH 7.0, 4 L), containing 5 mM dithiothreitol. The homogenate was filtered through four layers of cheesecloth into a beaker containing 10% (w/v) polyvinylpyrrolidone. The filtrate was centrifuged (12,000 x g, 15 min). The resulting supernatant was fractionated with (NH₄)₂SO₄, with proteins precipitating between 40 and 60% saturation recovered by centrifugation (10,000 x g, 1 h). The pellet was next reconstituted in a minimum amount of Tris-HCl buffer (20 mM, pH 8.0), containing 5 mM dithiothreitol (buffer A) and desalting using prepacked PD-10[®] columns (Sephadex G-25 medium) equilibrated with buffer A.

Affinity (Affi Blue Gel) Chromatography. The crude enzyme preparation (191 mg in buffer A, 5 nmol h⁻¹ mg⁻¹) was applied to an Affi Blue Gel column (2.6 x 70 cm) equilibrated in buffer A. After washing the column with 200 ml of buffer A, pinoresinol/lariciresinol reductase was eluted with a linear NaCl gradient

(1.5-5 M in 300 ml) in buffer A at a flow rate of 1 ml min⁻¹. Active fractions were stored (-80°C) until needed.

Hydrophobic Interaction Chromatography (Phenyl Sepharose). After thawing, ten preparations resulting from the Affi Blue chromatography step (150 mg, 51 nmol h⁻¹ mg⁻¹) were combined and applied to a Phenyl Sepharose column (1 x 10 cm) equilibrated in buffer A, containing 5 M NaCl. The column was washed with two bed volumes of the same buffer. Pinoresinol/lariciresinol reductase was eluted using a linear gradient of decreasing concentration of NaCl (5.0 M in 40 ml) in buffer A at a flow rate of 1 ml min⁻¹. Fractions catalyzing pinoresinol/lariciresinol reduction were combined and pooled.

Hydroxyapatite / Chromatography. Active protein (31 mg, 91 nmol h⁻¹ mg⁻¹) from the phenyl sepharose purification step was applied to an hydroxyapatite column (1.6 x 70 cm) equilibrated in 10 mM potassium phosphate buffer, pH 7.0, containing 5 mM dithiothreitol (buffer B). Pinoresinol/lariciresinol reductase was eluted with a linear gradient of potassium phosphate buffer, pH 7.0 (0.01-0.4 M in 200 ml) at a flow rate of 1 ml min⁻¹. Active fractions were combined. The buffer was then exchanged with buffer A using PD-10 prepakced columns.

Affinity (2',5'-ADP Sepharose) Chromatography. The enzyme solution resulting from the hydroxyapatite purification step (6.5 mg, 463 nmol h⁻¹ mg⁻¹) was next loaded on a 2',5'-ADP Sepharose (1 x 10 cm) column, previously equilibrated in buffer A containing 2.5 mM EDTA (buffer A') and then washed with 25 ml of buffer A'. Pinoresinol/lariciresinol reductase was eluted with a step gradient of NADP+ (0.3 mM in 10 ml) in buffer A' at a flow rate of 0.5 ml min⁻¹. [NADP+ (up to 3 mM) did not elute pinoresinol/lariciresinol reductase activity.] Because of the interference of the absorbance of the NADP+, it was not possible to directly monitor the eluent at 280 nm. Protein concentrations for each fraction were determined spectrophotometrically according to Bradford (Bradford, M.M., *Anal. Biochem.* 72:248-254 (1976)).

Hydroxyapatite II Chromatography. Fractions from the 2',5'-ADP Sepharose column that exhibited pinoresinol/lariciresinol reductase activity (0.85 mg, 1051 nmol h⁻¹ mg⁻¹) were combined and directly applied to a second hydroxyapatite column (1 x 3 cm), equilibrated in buffer B, with the enzyme eluted with a linear gradient of potassium phosphate buffer, pH 7.0 (0.01-0.4 M in 45 ml) at a flow rate of 1 ml min⁻¹.

Affinity (Affi Yellow) Chromatography - Active fractions (160 µg, 7960 nmol h⁻¹ mg⁻¹) from the second hydroxyapatite column purification step were next applied to a Reactive Yellow 3 Agarose column (1 x 3 cm), equilibrated in buffer A. Pinoresinol/lariciresinol reductase was eluted with a linear NaCl gradient (0.25 M in 100 ml) at a flow rate of 1 ml min⁻¹.

Fast Protein Liquid Chromatography (Superose 12 Chromatography) - Combined fractions from the Affi Yellow purification step having the highest activity (50 µg, 10,940 nmol h⁻¹ mg⁻¹) were pooled and concentrated to 1 ml, using a Centriflo 10 microconcentrator (Amicon, Inc.). The enzyme solution was then applied in portions of 200 µl to a fast protein liquid chromatography column (Superose 12, HR 10/30). Gel filtration was performed in a buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl and 5 mM dithiothreitol at a flow rate of 0.4 ml min⁻¹. Pinoresinol/lariciresinol reductase was eluted with 12.8 ml of the mobile phase. The active fractions which coincided with the UV profile (absorbance at 280 nm) were pooled (20 µg, 15,300 nmol h⁻¹ mg⁻¹) and desalting (PD-10 prepakced columns).

The foregoing purification protocol resulted in a 3060-fold purification of (+)-pinoresinol/(-)-lariciresinol reductase. As for many of the enzymes involved in phenylpropanoid metabolism, the protein was in very low abundance, i.e. 20 kg *F. intermedia* stems yielded only ~20 µg of the purified (+)-pinoresinol/(-)-lariciresinol reductase.

EXAMPLE 9

Characterization of Purified Pinoresinol/lariciresinol Reductases

from *Forsythia Intermedia*

Isoelectric Focusing and pI Determination. In all stages of the purification protocol, (+)-pinoresinol/(-)-lariciresinol reductase activities coeluted. Given this observation, it was essential to unambiguously ascertain whether more than one form of the protein existed, i.e., whether one form of the protein catalyzed the reduction of pinoresinol, and another form of the protein catalyzed the reduction of lariciresinol. To this end, the isoelectric point of pinoresinol/lariciresinol reductase was estimated by chromatofocusing on a MonoG HR 5/20 FPLC column.

Active fractions from the Superose 12 gel filtration column (Example 1) were pooled and the buffer exchanged with 25 mM Bis-Tris, pH 7.1, using prepacked PD-10 columns, equilibrated in the same buffer. The preparation so obtained was loaded on the chromatofocusing column and a pH gradient between 7.1 and 3.9 was

formed, using Polybuffer 74 as eluent at a flow rate of 0.5 ml min⁻¹. Aliquots (200 μ l) of each fraction were assayed for pinoresinol/lariciresinol reductase activities. The remainder of the fractions was used to determine the pH gradient.

Molecular Weight Determination. Application of the MonoP HR 5/20 FPLC column preparation of pinoresinol/lariciresinol reductase to SDS gradient gel electrophoresis (4-15% polyacrylamide) revealed the presence of two protein bands of similar apparent molecular weight, whose separation was achieved via anion-exchange chromatography on a MonoQ HR 5/5 FPLC matrix. Pooled fractions from the Sepharose 12 purification step (Example 1) were applied to a MonoQ HR 5/5 column (Pharmacia), equilibrated in buffer A. The column was washed with 10 ml of buffer A and pinoresinol/lariciresinol reductase activity eluted using a linear NaCl gradient (0-500 mM in 50 ml) in buffer A at a flow rate of 0.5 ml min⁻¹. Aliquots (30 μ l) of the collected fractions were analyzed by SDS-polyacrylamide gel electrophoresis, using a gradient (4-15% acrylamide) gel. Proteins were visualized by silver staining. Active fractions 34 through 37 (27,760 nmol h⁻¹ mg⁻¹) and 38 through 41 (30,790 nmol h⁻¹ mg⁻¹) were pooled separately and immediately used for characterization.

The two protein bands thus resolved under "denaturing" conditions had apparent molecular masses of ~36 and ~35 kDa, respectively. Each of the two reductase forms had a pI of 5.7.

Native molecular weights of each reductase isoform were estimated via comparison of their elution behavior on Superose 12, Superose 6 and Superdex 75 gel filtration FPLC columns with the elution behavior of calibrated molecular weight standards. Gel filtration was carried out as set forth in Example 8. For each reductase, an apparent native molecular weight of 59,000 was calculated based on its elution volume, in contrast to that of ~36,000 and ~35,000 by SDS-polyacrylamide gel electrophoresis. While the discrepancy between molecular weights from gel filtration and SDS-PAGE remains unknown, it can "tentatively" be proposed that although the native protein likely exists as a dimer, it could also be a monomer of asymmetric shape, thereby altering its effective Stokes radius (Cantor, C.R., and Shimmel, P.R., *Biophysical Chemistry*, Part II, W.H. Freeman and Company, San Francisco, CA (1980); Stellwagen, E., *Methods in Enzymology* 182:317-328 (1990); as reported for human thioredoxin reductase (Oblong, J.E., et al., *Biochemistry* 32:7271-7277 (1993)) and yeast metalloendopeptidase (Hrycyna, C.A., and Clarke, S., *Biochemistry* 32:11293-11301 (1993)).

pH and Temperature Optima. To determine the pH-optimum of pinoresinol/lariciresinol reductase, the enzyme preparation from the gel Superose 12 filtration step (Example 8) was assayed utilizing standard assay conditions (Example 8), except that the buffer was replaced with 50 mM Bis-Tris Propane buffer in the pH range of 6.3 to 9.4. The pH optimum was found to be pH 7.4.

The temperature optimum of pinoresinol/lariciresinol reductase was examined in the range between 4°C and 80°C under standard assay conditions (Example 8) utilizing the enzyme preparation from the gel filtration step (Example 8). At optimum pH, the temperature optimum for the reductase activity was established to be ~30°C.

Kinetic Parameters. Velocity studies were carried out to ascertain whether the two reductase isoforms catalyzed distinct reductions; i.e., that of the conversion of (+)-pinoresinol to (+)-lariciresinol, and (+)-lariciresinol to (-)-secosolaresinol, respectively, or whether either displayed a preference for (+)-pinoresinol or (+)-lariciresinol as substrates. The initial velocity studies were carried out individually utilizing the two isoforms of the enzyme, and individually employing both (+)-pinoresinol and (+)-lariciresinol as substrates. Initial velocity studies were performed in triplicate experiments, using 50 mM Bis-Tris Propane buffer, pH 7.4 containing 5 mM dithiothreitol, pure enzyme (after MonoQ anion-exchange chromatography), ten different substrate concentrations (between 8.8 and 160 μ M) at a constant NADPH concentration (80 μ M). Incubations were carried out at 30 °C for 10 min (within the linear kinetic range). Kinetic parameters were determined from Lineweaver-Burk plots.

Importantly, the kinetic parameters were essentially the same for both the 35 kDa and the 36 kDa forms of the enzyme (i.e., Km for pinoresinol: 27±1.5 μ M for the 35 kDa form of the enzyme, and 23±1.3 μ M for the 36 kDa form of the enzyme; Km for lariciresinol: 121±5.0 μ M for the 35 kDa form of the enzyme and 123±6.0 μ M for the 36 kDa form of the enzyme). In an analogous manner, apparent maximum velocities (expressed as μ mol h⁻¹ mg⁻¹ of protein) were also essentially identical (i.e., Vmax for pinoresinol: 16.2±0.4 for the 35 kDa form of the enzyme and 17.3±0.5 for the 36 kDa form of the enzyme; for lariciresinol: 25.2±0.7 for the 35 kDa form of the enzyme and 29.9±0.7 for the 36 kDa form of the enzyme). Thus, all available evidence suggests that (+)-pinoresinol/(-)-lariciresinol reductase exists as two isoforms, with each capable of catalyzing the reduction of both substrates. How this reduction is carried out, i.e., whether both reductions are done in tandem, in either

quinone or furano ring form, awaits further study using a more abundant protein source.

Enzymatic Formation of (+)-[7R,2H]Lariciresinol. Since the two (+)-pinoresinol/(+)-lарицирсинол reductase isoforms exhibited essentially identical catalytic characteristics, the Sepharose 12 enzyme preparation (Example 8), containing both isoforms, was used to examine the stereospecificity of the hydride transfer. The strategy adopted utilized selective deuteration labeling using NADP²H as cofactor for the reduction of (+)-pinoresinol, with the enzymatic product, (+)-lарицирсинол, being analyzed by ¹H NMR and mass spectroscopy. Thus, a solution of (+)-pinoresinols (5.2 mM in MeOH, 4 ml) was added to Tris-HCl buffer (20 mM, pH 8.0, containing 5 mM dithiothreitol, 22 ml) and stereospecifically deuterio-labeled [4R,2²H]NADPH (20 mM in H₂O, 4 ml) prepared via the method of Anderson and Lin (Anderson, J.A., and Lin B.K., *Phytochemistry* 32:811-812 (1993)), with the whole added to the enzyme preparation (20 ml). After incubation at 30°C for 1h with shaking, the assay mixture was extracted with EtOAc (2 x 50 ml). The EtOAc soluble fraction was combined, washed with saturated NaCl (50 ml), dried (Na₂SO₄), and evaporated to dryness in vacuo. The resulting extract was reconstituted in a minimum amount of EtOAc, applied to a silica gel column (0.5 x 7 cm), and eluted with EtOAc/hexanes (1:2). Fractions containing the enzymatic product were combined and evaporated to dryness.

The enzymatic product was established to be (+)-[7²R,2H]lарицирсинол, as evidenced by the disappearance of the 7'-proR proton at δ 2.51 ppm due to its replacement by deuterium and by its molecular ion at (m/z) 361 (M⁺⁺1) corresponding to the presence of one deuterium atom at C-7. ¹H NMR (300 MHz, CDCl₃): 2.39 (m, H, C8H), 2.71 (m, H, C8' H), 2.88 (δ, ¹H, C7S, δ=5.0 Hz, C7HS), 3.73 (δδ, ¹H, J^{8,9}a=7.0 Hz, J^{9,10}a=8.5 Hz, C9H), 3.76 (δδ, ¹H, J^{8,9}S=6.5 Hz, J^{9,10}S=8.5 Hz, C9HS), 3.86 (s, ³H, OCH₃), 3.88 (s, ³H, OCH₃), 3.92 (δδ, ¹H, J^{8,9}R=6.0 Hz, J^{9,R}, J^{9,S}=9.5 Hz, C9HR), 4.04 (δδ, ¹H, J^{8,9}a=7.0 Hz, J^{9,10}a=8.5 Hz, C9Ha), 4.77 (δ, ¹H, J^{7,8}=6.5 Hz, C7H), 6.68 - 6.70 (m, H, ArH), 6.75 - 6.85 (m, 4H, ArH); MS m/z (%): 361 (M⁺⁺1, 71.2), 360 (M⁺, 31.1), 237 (11.1), 153 (41.5), 152 (20.2), 151 (67.0), 138 (100), 137 (71.1).

Thus, hydride transfer from (+)-pinoresinol to (+)-lарицирсинол had occurred in a manner whereby only the 7'-proR hydrogen position of (+)-lарицирсинол was deuterated. An analogous result was observed for the conversion of (+)-lарицирсинол

into (+)-secoisolariciresinol, thereby establishing that the overall hydride transfer was completely stereospecific.

EXAMPLE 10

Amino Acid Sequence Analysis of Purified Pinoresinol/Lariciresinol Reductase from *Forsythia intermedia*

Pinoresinol/Lariciresinol Reductase Amino Acid Sequencing. The (+)-pinoresinol/(+)-lарицирсинол reductase N-terminal amino acid sequence was obtained from each of the purified proteins, and a mixture of both, using an Applied Biosystems protein sequencer with on-line HPLC detection. The N-terminal sequence was the same for both isoforms (SEQ ID No:36).

For trypsin digestion, 150 pmol of the enzyme purified from the Sepharose 12 column (Example 8) was suspended in 0.1 M Tris-HCl (50 μl, pH 8.5), with urea added to give a final concentration of 8 M in 77.5 μl. The mixture was incubated for 15 min at 50°C, then 100 mM iodoacetamide (2.5 μl) was added, with the whole kept at room temperature for 15 min. Trypsin (1 μg in 20 μl) was then added, with the mixture digested for 24 h at 37°C, after which TFA (4 μl) was added to stop the enzymatic reaction.

The resulting mixture was subjected to reversed phase HPLC analysis (C-8 column, Applied Biosystems), this being eluted with a linear gradient over 2 h from 0 to 100% acetonitrile (in 0.1% TFA) at a flow rate of 0.2 ml/min with detection at 280 nm. Fractions containing individual oligopeptide peaks were collected manually and directly submitted to amino acid sequencing. Four tryptic fragments were resolved in sufficient quantity to permit amino acid sequence determination. (SEQ ID Nos:37-40).

Cyanogen bromide digestion was performed by incubation of 1.50 pmol of the reductase purified from the Sepharose 12 column (Example 8) with 0.5 M cyanogen bromide in 70% formic acid for 40 h at 37°C, following which the cyanogen bromide and formic acid were removed by centrifugation under reduced pressure (SpeedVac). The resulting oligopeptide fragments were separated by HPLC and three were resolved in sufficient quantity to permit sequencing (SEQ ID Nos:41-43).

EXAMPLE 11

Cloning of Pinoresinol/Lariciresinol Reductase from *Forsythia intermedia*

Plant Materials. *Forsythia intermedia* plants were either obtained from Bailey's Nursery (var. Lynwood Gold, Sl, Paul, MN), and maintained in Washington State University greenhouse facilities, or were gifts from the local community.

Materials. All solvents and chemicals used were reagent or HPLC grade.

UV RNA and DNA determinations at OD₂₆₀ were obtained on a Lambda 6 UV/VIS spectrophotometer. A Temptron II thermocycler (Thermolyne) was used for all PCR amplifications. Tag thermostable DNA polymerase was obtained from Promega, whereas restriction enzymes were from Gibco BRL (HaeII), Boehringer Mannheim (Sau3A) and Promega (TaqI). pT7Blue T-vector and competent NovaBlue cells were purchased from Novagen and radiolabeled nucleotides [α -³²P]dCTP and [γ -³²P]ATP were from DuPont NEN.

Oligonucleotide primers for polymerase chain reaction (PCR) and sequencing were synthesized by Gibco BRL Life Technologies GENECLEAN II® kits (BIO 101 Inc.) were used for purification of PCR fragments, with the gel-purified DNA concentrations determined by comparison to a low DNA mass ladder (Gibco BRL) in 1.5% agarose gels.

Forsythia RNA Isolation. Initial attempts to isolate functional *F. intermedia* RNA from fast-growing, green stem tissue were unsuccessful, due to difficulties encountered via facile oxidation by its plant phenolic constituents. This problem was, however, successfully overcome by utilization of an RNA isolation procedure, specifically designed for woody plant tissue, which uses low pH, and reducing conditions in the extraction buffer to prevent oxidation (Dong, Z.D., and Dunstan, D.I., *Plant Cell Reports* 15: 516-521 (1996)).

Forsythia intermedia stem cDNA Library Synthesis. Total RNA (~300 μ g/g fresh weight) was obtained from young green stems of greenhouse-grown *Forsythia intermedia* plants (var. Lymwood Gold) (Dong, Z.D. and Dunstan, D.I., *Plant Cell Reports* 15: 516-521 (1996)). A *Forsythia intermedia* stem cDNA library was constructed using 5 μ g of purified poly A+ mRNA (Oligotex-dT₃₀ Suspension, QIAGEN) with the ZAP-cDNA® synthesis kit, the Uni-ZAP™ XR vector and the Gigapack® II Gold packaging extract (Stratagene), with a titer of 1.2 x 10⁶ PFU for the primary library. A portion (30 ml) of the amplified library (1.2 x 10¹⁰ PFU/ml; 158 ml total) was used to obtain pure cDNA library DNA for PCR.

(Sambrook, J. et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994); Ausubel, F.M. et al., *Current Protocols in Molecular Biology*, 2 volumes, Greene Publishing Associates and Wiley-Interscience, John Wiley & Sons NY (1991)).

Phenol/mL/Laricresinol Reductase DNA Probe Synthesis - The N-terminal and internal peptide amino acid sequences were used to construct the degenerate

oligonucleotide primers. Specifically, the primer PLRNS (SEQ ID No:44) was based on the sequence of amino acids 7 to 13 of the N-terminal peptide (SEQ ID No:36). The primer PLR14R (SEQ ID No:45) was based on the sequence of amino acids 2 to 8 of the internal peptide sequence set forth in (SEQ ID No:37). The primer PLR15R (SEQ ID No:46) was based on the sequence of amino acids 9 to 15 of the internal peptide sequence set forth in (SEQ ID No:37). The sequence of amino acids 9-to 15 of the internal peptide sequence set forth in SEQ ID No:37, upon which the sequence of primer PLR15R (SEQ ID No:46) was based, also corresponded to the sequence of primer PLR15R (SEQ ID No:46) was based, also corresponded to the sequence of amino acids 4 to 10 of the cyanogen bromide-generated, internal NovaBlue cells were purchased from Novagen and radiolabeled nucleotides [α -³²P]dCTP and [γ -³²P]ATP were from DuPont NEN.

Oligonucleotide primers for polymerase chain reaction (PCR) and sequencing were synthesized by Gibco BRL Life Technologies GENECLEAN II® kits (BIO 101 Inc.) were used for purification of PCR fragments, with the gel-purified DNA concentrations determined by comparison to a low DNA mass ladder (Gibco BRL) in 1.5% agarose gels.

Forsythia RNA Isolation. Initial attempts to isolate functional *F. intermedia* RNA from fast-growing, green stem tissue were unsuccessful, due to difficulties encountered via facile oxidation by its plant phenolic constituents. This problem was, however, successfully overcome by utilization of an RNA isolation procedure, specifically designed for woody plant tissue, which uses low pH, and reducing conditions in the extraction buffer to prevent oxidation (Dong, Z.D., and Dunstan, D.I., *Plant Cell Reports* 15: 516-521 (1996)).

Forsythia intermedia stem cDNA Library Synthesis. Total RNA (~300 μ g/g fresh weight) was obtained from young green stems of greenhouse-grown *Forsythia intermedia* plants (var. Lymwood Gold) (Dong, Z.D. and Dunstan, D.I., *Plant Cell Reports* 15: 516-521 (1996)). A *Forsythia intermedia* stem cDNA library was constructed using 5 μ g of purified poly A+ mRNA (Oligotex-dT₃₀ Suspension, QIAGEN) with the ZAP-cDNA® synthesis kit, the Uni-ZAP™ XR vector and the Gigapack® II Gold packaging extract (Stratagene), with a titer of 1.2 x 10⁶ PFU for the primary library. A portion (30 ml) of the amplified library (1.2 x 10¹⁰ PFU/ml; 158 ml total) was used to obtain pure cDNA library DNA for PCR.

Associates and Wiley-Interscience, John Wiley & Sons NY (1991)).

Phenol/mL/Laricresinol Reductase DNA Probe Synthesis - The N-terminal and internal peptide amino acid sequences were used to construct the degenerate

oligonucleotide primers. Specifically, the primer PLRNS (SEQ ID No:44) was based on the sequence of amino acids 7 to 13 of the N-terminal peptide (SEQ ID No:36). The primer PLR14R (SEQ ID No:45) was based on the sequence of amino acids 2 to 8 of the internal peptide sequence set forth in (SEQ ID No:37). The primer PLR15R (SEQ ID No:46) was based on the sequence of amino acids 9 to 15 of the internal peptide sequence set forth in (SEQ ID No:37). The sequence of amino acids 9-to 15 of the internal peptide sequence set forth in SEQ ID No:37, upon which the sequence of primer PLR15R (SEQ ID No:46) was based, also corresponded to the sequence of primer PLR15R (SEQ ID No:46) was based, also corresponded to the sequence of amino acids 4 to 10 of the cyanogen bromide-generated, internal NovaBlue cells were purchased from Novagen and radiolabeled nucleotides [α -³²P]dCTP and [γ -³²P]ATP were from DuPont NEN.

Oligonucleotide primers for polymerase chain reaction (PCR) and sequencing were synthesized by Gibco BRL Life Technologies GENECLEAN II® kits (BIO 101 Inc.) were used for purification of PCR fragments, with the gel-purified DNA concentrations determined by comparison to a low DNA mass ladder (Gibco BRL) in 1.5% agarose gels.

Forsythia RNA Isolation. Initial attempts to isolate functional *F. intermedia* RNA from fast-growing, green stem tissue were unsuccessful, due to difficulties encountered via facile oxidation by its plant phenolic constituents. This problem was, however, successfully overcome by utilization of an RNA isolation procedure, specifically designed for woody plant tissue, which uses low pH, and reducing conditions in the extraction buffer to prevent oxidation (Dong, Z.D., and Dunstan, D.I., *Plant Cell Reports* 15: 516-521 (1996)).

Forsythia intermedia stem cDNA Library Synthesis. Total RNA (~300 μ g/g fresh weight) was obtained from young green stems of greenhouse-grown *Forsythia intermedia* plants (var. Lymwood Gold) (Dong, Z.D. and Dunstan, D.I., *Plant Cell Reports* 15: 516-521 (1996)). A *Forsythia intermedia* stem cDNA library was constructed using 5 μ g of purified poly A+ mRNA (Oligotex-dT₃₀ Suspension, QIAGEN) with the ZAP-cDNA® synthesis kit, the Uni-ZAP™ XR vector and the Gigapack® II Gold packaging extract (Stratagene), with a titer of 1.2 x 10⁶ PFU for the primary library. A portion (30 ml) of the amplified library (1.2 x 10¹⁰ PFU/ml; 158 ml total) was used to obtain pure cDNA library DNA for PCR.

Associates and Wiley-Interscience, John Wiley & Sons NY (1991)).

Phenol/mL/Laricresinol Reductase DNA Probe Synthesis - The N-terminal and internal peptide amino acid sequences were used to construct the degenerate

-55-

(SEQ ID No:74) and U19mer (SEQ ID No:75) primers according to the manufacturer's (Novagen's) instructions.

Restriction analysis was performed to determine whether all inserts for each combination of primers and template were the same. Restriction analysis was carried out as follows: each of the inserts was amplified by PCR utilizing the R20 (SEQ ID No:74) and U19 (SEQ ID No:75) primers. To 20 μ l each of a 100 μ l PCR reaction were added 4 units HaeII, 1.5 units Sau3A or 5 units TaqI restriction enzyme. Restriction digestions were allowed to proceed for 60 min at 37°C for HaeII and Sau3A and at 65°C for TaqI reactions. Restriction products were resolved in 1.5% agarose gels giving one restriction group for all inserts tested.

Five of the resulting, recombinant plasmids were selected for DNA sequencing. The inserts from three of the recombinant plasmids (called pT7PLR1-pT7PLR3) were generated by a combination of primers PLRNS (SEQ ID No:44) and PLR5R (SEQ ID No:46) with the 400 bp PCR product as substrate. The inserts from the remaining two recombinant plasmids (called pT7PLR4 and pT7PLR5) were generated from a combination of primers PLRNS (SEQ ID No:44) and PLR4R (SEQ ID No:45) and the 380 bp PCR product as substrate. All of the five, sequenced PCR products contained the same open reading frame.

The (+)-pinoresinol/(-)-lарicresinol reductase probe was constructed as follows: five, 100 μ l PCR reactions were performed as described above with 10 ng pT7PLR3 DNA with primers PLRNS (SEQ ID No:44) and PLR5R (SEQ ID No:46). Gel-purified pT7PLR3 cDNA insert (50 ng) was used with Pharmacia's T7QuickPrime® Kit and [α -³²P]dCTP, according to kit instructions, to produce a radiolabelled probe (in 0.1 ml), which was purified over BioSpin 6 columns (Bio-Rad) and added to carrier DNA (0.9 ml of 0.5 mg/ml sheared salmon sperm DNA obtained from Sigma).

Library Screening. 600,000 PFU of *F. intermedia* amplified cDNA library were plated for primary screening, according to Stratagene's instructions. Plaques were blotted onto Magna Nylon membrane circles (Milton Separations Inc.), which were then allowed to air dry. The membranes were placed between two layers of Whatman® 3MM Chr paper, cDNA library phage DNA was fixed to the membranes and denatured in one step by autoclaving for 2 min at 100°C with fast exhaust. The membranes were washed for 30 min at 37°C in 6X standard saline citrate (SSC) and 0.1% SDS and prehybridized for 5 h with gentle shaking at 57-58°C in preheated 6X

SSC, 0.5% SDS and 5X Denhardt's reagent (hybridization solution, 300 ml) in a crystallization dish (190x75 mm).

The [³²P]radiolabeled probe was denatured (boiling, 10 min), quickly cooled (ice, 15 min) and added to a preheated fresh hybridization solution (60 ml, 58°C) in a crystallization dish (150x75 mm). The prehybridized membranes were next added to this dish, which was then covered with plastic wrap. Hybridization was performed for 18 h at 57-58°C with gentle shaking. The membranes were washed in 4X SSC and 0.5% SDS for 5 min at room temperature, transferred to 2X SSC and 0.5% SDS (at room temperature) and incubated at 57-58°C for 20 min with gentle shaking, wrapped with plastic wrap to prevent drying and finally exposed to Kodak X-OMAT AR film for 24 h at -80°C with intensifying screens.

This screening procedure resulted in more than 350 positive plaques, with twenty (of different signal intensities) being subjected to two additional rounds of screening. After final purification, six of the twenty cDNAs were subcloned by *in vivo* excision into pBluescript. These six cDNAs were called pfr-Fil to pfr-F16 (SEQ ID Nos:47, 49, 51, 53, 55, 57).

In vivo Excision and Sequencing of pfr-Fil-pfr-F16 Phagemids. The six purified cDNA clones were rescued from the phage following Stratagene's *in vivo* excision protocol. Both strands of the six different cDNAs (pfr-Fil to pfr-F16) that coded for (+)-pinoresinol/(-)-lарicresinol reductase were completely sequenced using overlapping sequencing primers.

Purification of DNA for sequencing employed a QIAwell Plus plasmid purification system (QIAGEN) followed by PEG precipitation (Sambrook, J., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)), with DNA sequences determined using an Applied Biosystems' Model 373A automated sequencer. DNA and amino acid sequence analyses were performed using the Unix-based GCG Wisconsin Package (Program Manual for the Wisconsin Package, Version 8, September 1994, *Genetics Computer Group*, 575 Science Drive, Madison, Wisconsin, USA 53711; Rice, P., 30 *Program Manual for the EGCG Package*, Peter Rice, The Sanger Centre, Hinxton Hall, Cambridge, CB10 1RQ, England (1996) and the ExPASy World Wide Web molecular biology server (Geneva University Hospital and University of Geneva, Geneva, Switzerland).

All six cDNAs had the same coding but different 5'-untranslated regions. On the other hand, analysis of the 3'-untranslated region of each of the six cDNAs

established that all were truncated versions of the longest cDNAs' 3'-region. Preliminary RNA gel blot analysis with total RNA from greenhouse-grown plant stem tips confirmed a single transcript with a length of approximately 1.2 kb.

RNA gel blot analysis. For RNA gel blot analysis, total RNA (30 µg per lane) from *F. intermedia* stem tips was separated by size by denaturing agarose gel electrophoresis. The RNA was transferred to charged nylon membranes (GeneScreen Plus®, Dupont NEN), cross-linked to the membrane (Stratalinker from Stratagene), prehybridized, hybridized with the same probe used to screen the cDNA library during cDNA cloning and washed according to the manufacturer's instructions for aqueous hybridization conditions. The membrane was then exposed to Kodak X-OMAT film for 48 hr at -80°C with intensifying screens.

EXAMPLE 12

Expression of (+)-Pinoresinol/(-)-Laricresinol Reductase cDNA pfr-F1 in *E. coli*

Expression in Escherichia coli. In order to confirm that the putative (+)-pinoresinol/(-)-laricresinol reductase cDNAs encoded functional (+)-pinoresinol/(-)-laricresinol reductase, the cDNAs putatively encoding (+)-pinoresinol/(-)-laricresinol reductase were heterologously expressed in *E. coli*. Heterologous expression was also necessary in order to obtain sufficient protein to enable the systematic study of the precise biochemical mechanism of (+)-pinoresinol/(-)-laricresinol reductase at a future date.

Examination of the six putative (+)-pinoresinol/(-)-laricresinol reductase clones revealed that one, pfr-F1 (SEQ ID No:47), was in frame with the α -complementation particle of β -galactosidase in pBluescript. This was fortuitous, since it potentially provided a facile means to express the fully functional fusion protein, and hence to provide proof that the cloned sequence was correct.

Purified plasmid DNA from pfr-F1 (SEQ ID No:47) was transformed into NovaBlue cells according to Novagen's instructions. Transformed cells (5 ml cultures) were grown at 37°C with shaking (225 rpm) to mid log phase (OD₆₀₀ = 0.5) in LB medium (Sambrook, J., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)) supplemented with 12.5 µg ml⁻¹ tetracycline and 50 µg ml⁻¹ ampicillin. IPTG (isopropyl β -D-thiogalactopyranoside) was then added to a final concentration of 10 mM, and the cells were allowed to grow for 2 h. Cells were collected by centrifugation and resuspended in 500 µl (per 5 ml culture tube) buffer (20 mM Tris-HCl, pH 8.0, 5 mM dithiothreitol). Lysozyme (5 µl of 0.1 mg ml⁻¹, Research

Organics, Inc.) was next added and following incubation for 10 min, the cells were lysed by sonication (3 x 15 s). After centrifugation at 14,000 x g at 4°C for 10 min, the supernatant was removed and assayed for (+)-pinoresinol/(-)-laricresinol reductase activity (21.0 µl supernatant per assay) as described in Example 8.

Catalytic activity was established by incubating cell-free extracts for 2 h at 30°C with (+)-pinoresinols (0.4 mM) and [A^3H]-NADPH (0.8 mM) under standard conditions. Following incubation, unlabeled (-)-laricresinols and (\pm)-secoisolaricresinols were added as radiochemical carriers, with each lignan isolated by reversed-phase HPLC. Controls included assays of a pinoresinol/laricresinol reductase cDNA which contains an out-of-frame cDNA insert, with all assay components, as well as pfr-F1 (SEQ ID No:47) and an out-of-frame pinoresinol/laricresinol reductase cDNA with no substrate except [A^3H]-NADPH. Separation of products and chiral identification were performed by HPLC as previously described (Ohi, A., et al., *J. Biol. Chem.* 268:27026-27033 (1993)).

Subsequent chiral HPLC analysis revealed that both (-)-laricresinol and (-)-secoisolaricresinol, but not the corresponding antipodes, were radiolabeled (total activity: 54 nmol h⁻¹ mg⁻¹). By contrast, no catalytic activity was detected either in the absence of (+)-pinoresinols, or when control cells were used which contained a plasmid in which the cDNA insert was not in-frame with the β -galactosidase gene.

Thus, the heterologously expressed (+)-pinoresinol/(-)-laricresinol reductase and the plant protein function in precisely the same enantiospecific manner.

EXAMPLE 13

Sequence and Homology Analysis of the cDNA Insert of Clone pfr-F1 (SEQ ID No:47) Encoding (+)-Pinoresinol/(-)-Laricresinol Reductase

Sequence Analysis. The full length sequence of the cloned (+)-pinoresinol/(-)-laricresinol reductase pfr-F1 (SEQ ID No:47) contained all of the peptide sequences determined by Edman degradation of digest fragments.

The single ORF predicts a polypeptide of 312 amino acids (SEQ ID No:48) with a calculated molecular mass of 34.9 kDa, in close agreement with the value (~35 or ~36 kDa) estimated previously by SDS-PAGE for the two isoforms of (+)-pinoresinol/(-)-laricresinol reductase. An equal number of acidic and basic residues are also present, with a theoretical isoelectric point (pI) of 7.09, in contrast to that experimentally obtained by chromatofocusing (pI ~5.7).

The amino acid composition reveals seven methionine residues. Interestingly, the N-terminus of the plant-purified enzyme lacks the initial

5 10 15 20 25 30 35

-59-

methionine, this being the most common post-translational protein modification known. Consequently, the first methionine in the cDNA can be considered to be the site of translational initiation. The sequence analysis also reveals a possible N-glycosylation site at residue 215 (although no secretory targeting signal is present), and seven possible protein phosphorylation sites at residues 50 and 228 (protein kinase C-type), residues 228, 250, 302 and 303 (cascin kinase II-type) and residue 301 (tyrosine kinase type).

Regions of the pinoresinol/lariciresinol polypeptide chain (SEQ ID NO:48) were also identified that contained conserved sequences associated with NADPH binding (Jornvall, H., in *Dehydrogenases Requiring Nicotinamide Coenzymes* (Jeffery, J., ed) pp. 126-148, Birkhäuser Verlag, Basel (1980); Branden, C., and Tooze, J., *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York and London (1991); Wierenga, R.K. et al., *J. Mol. Biol.* 187:101-108 (1986)). There is a limited number of invariant amino acids in the sequences of different reductases which are viewed as indicative of NADPH binding sites. These include three conserved glycine residues with the sequence G-X-G-X-X-G (SEQ ID No:76), where X is any residue, and six conserved hydrophobic residues. The glycine-rich region is considered to play a central role in positioning the NADPH in its correct conformation. In this regard, a comparison of the N-terminal region of (+)-pinoresinol/(-)-lariciresinol reductase with that of the conserved, NADPH-binding regions of *Drosophila melanogaster* alcohol dehydrogenase (Branden, C., and Tooze, J., *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York and London (1991)), *Pinus taeda* cinnamyl alcohol dehydrogenase (MacKay J.J. et al., *Mol. Gen. Genet.* 247:537-545 (1995)), dogfish muscle lactate dehydrogenase (Branden, C., and Tooze, J., *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York and London (1991)) and human erythrocyte glutathione reductase (Branden, C., and Tooze, J., *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York and London (1991)), revealed some interesting parallels. The invariant glycine residues are aligned in every case, as are four of the six hydrophobic residues required for the correct packaging in the formation of the domain. Hence, the NADPH-binding site of (+)-pinoresinol/(-)-lariciresinol reductase isoforms is localized close to the N-terminus.

Homology Analysis: Comparison to Isoflavone Reductase. A BLAST search (Altschul, S.F. et al., *J. Mol. Biol.* 215:403-410 (1990)) was conducted with the

-60-

translated amino acid sequence of (+)-pinoresinol/(-)-lariciresinol reductase (SEQ ID No:48) against the non-redundant peptide database at the National Center for Biotechnology Information. Significant homology was noted for (+)-pinoresinol/(-)-lariciresinol reductase with various isoflavone reductases from the legumes, *Cicer arietinum* (Tiemann, K., et al., *Eur. J. Biochem.* 200:751-757 (1991)) (63.5% similarity, 44.4% identity), *Medicago sativa* (Paiva, N.L., et al., *Plant Mol. Biol.* 17:653-667 (1991)) (62.6% similarity, 42.0% identity) and *Pinus sativum* (Paiva, N.L., et al., *Arch. Biochem. Biophys.* 312:501-510 (1994)) (61.6% similarity, 41.3% identity). This observation is of considerable interest since isoflavonoids are formed via a related branch of phenylpropanoid-acetate pathway metabolism. Specifically, isoflavone reductases catalyze the reduction of α,β -unsaturated ketones during isoflavonoid formation. For example, the *Medicago sativa* L. isoflavone reductase catalyzes the stereospecific conversion of 2'-hydroxyformononetin to (3R)-vestitone in the biosynthesis of the phytalexin, (2 γ)-medicarpin (Paiva, N.L. et al., *Plant Mol. Biol.* 17:653-667 (1991)). This sequence similarity may be significant given that both lignans and isoflavonoids are offshoots of general phenylpropanoid metabolism, with comparable plant defense functions and pharmacological roles, e.g., as "phytoestrogens". Consequently, since both reductases catalyze very similar reactions, it is tempting to speculate that the isoflavone reductases may have evolved from (+)-pinoresinol/(-)-lariciresinol reductase. This is considered likely since the lignans are present in the pieridophytes, hornworts, gymnosperms and angiosperms; hence their pathways apparently evolved prior to the isoflavonoids (Gang et al., In *Phytochemicals for Pest Control*, Hedin et al., eds, ACS Symposium Series, Washington D.C., 658:58-59 (1997)).

Comparable homology was also observed with putative isoflavone reductase "homologs" from *Arabidopsis thaliana* (Babiychuk, E., et al., Direct Submission (25-MAY-1995) to the EMBL/GenBank/DDBJ databases (1995)) (65.9% similarity, 50.8% identity), *Nicotiana tabacum* (Hibi, N., et al., *Plant Cell* 6:723-735 (1994)) (64.6% similarity, 47.2% identity), *Solanum tuberosum* (van Eldik, G.J., et al., (1995) Direct submission (06-OCT-1995) to the EMBL/GenBank/DDBJ databases) (65.5% similarity, 47.7% identity) *Zea mays* (Petrucio, S., et al., *Plant Cell* 8:69-80 (1996)) (61.6% similarity, 44.9% identity) and especially *Lupinus albus* (Attuci, S., et al., Personal communication and direct submission (06/09/96) to the EMBL/Genbank/DDBJ databases (1996)) (85.9% similarity, 66.2% identity).

By contrast, homology with other NADPH-dependent reductases was significantly lower: for example, dihydrafavonol reductases from *Petunia hybrida* (Beld, M., et al., *Plant Mol. Biol.* 13:491-502 (1989)) (43.2% similarity, 21.5% identity) and *Hordeum vulgare* (Kristiansen, K.N., and Rohde, W., *Mol. Gen. Genet.* 230:49-59 (1991)) (46.2% similarity, 21.1% identity), chalcone reductase from *Medicago sativa* (Ballance, G.M., and Dixon, R.A., *Plant Physiol.* 107:1027-1028 (1995)) (39.3% similarity, 15.8% identity), chalcone reductase "homolog" from *Sesbania rostrata* (Goomachiig, S., et al., (1995) Direct Submission (13-MAR-1995) to the EMBL/CeMB/DBJ databases) (47.6% similarity, 24.1% identity), 10 cholesterol dehydrogenase from *Nocardioides sp.* (Honouchi, S., et al., *Appl. Environ. Microbiol.* 57:1306-1303 (1991)) (46.6% similarity, 21.0% identity) and 3-hydroxy-5-ene steroid dehydrogenase from *Rattus norvegicus* (Zhao, H.-F., et al., *Journal Endocrinology* 127:327-329 (1990)) (43.5% similarity, 20.6% identity). Thus, sequence analysis establishes significant homology between 15 (+)-pinoresinol/(-)-lарciresinol reductase, isoflavone reductases and putative isoflavone reductase "homologs" which do not possess isoflavone reductase activity.

EXAMPLE 14

cDNA Cloning of *Thuya plicata* (-)-Pinoresinol/(-)-Lарciresinol Reductases

Plant Materials. Western red cedar plants (*Thuya plicata*) were maintained in 20 Washington State University greenhouse facilities.

Materials. All solvents and chemicals used were reagent or HPLC grade. *Tag* thermostable DNA polymerase and restriction enzymes (*Sac*' and *Xba*'I) were obtained from Promega. pT7Blue T-vector and competent NovaBlue cells were purchased from Novagen and radiolabeled nucleotide [α -³²P]dCTP was purchased from DuPont NEN.

Oligonucleotide primers for polymerase chain reaction (PCR) and sequencing were synthesized by Gibco BRL Life Technologies. GENECLEAN II[®] kits (BIO101 Inc.) were used for purification of PCR fragments with the gel-purified DNA concentrations determined by comparison to a low DNA mass ladder (Gibco BRL) in 1.3% agarose gels.

Instrumentation. UV (including RNA and DNA determinations at OD₂₆₀) spectra were recorded on a Lambda 6 UV/VIS spectrophotometer. A Tempronic II thermocycler (Thermolyne) was used for all PCR amplifications. Purification of plasmid DNA for sequencing employed a QIAwell Plus plasmid purification system (Qiagen) followed by PEG precipitation (Sambrook, J., et al., *Molecular Cloning*: A

Laboratory Manual, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)) or Wizard[®] Plus SV Minipreps DNA Purification System (Promega), with DNA sequences determined using an Applied Biosystems Model 373A automated sequencer.

5 *Thuya plicata* cDNA Library Synthesis. Total RNA (6.7 μ g/g fresh weight) was obtained from young green leaves (including stems) of greenhouse-grown western red cedar plants (*Thuya plicata*) according to the method of Lewinsohn et al. (Lewinsohn, E., et al., *Plant Mol. Biol. Rep.* 12:20-25 (1994)). A *T. plicata* cDNA library was constructed using 3 μ g of purified poly(A)⁺ mRNA (Oligotex-dT[™] Suspension, Qiagen) with the ZAP-cDNA[®] synthesis kit, the Uni ZAP[™] XR vector, and the Gigapack[®] II Gold packaging extract (Stratagene), with a titer of 1.2 \times 10⁵ pfu for the primary library. The amplified library (7.1 \times 10⁸ pfu/ml; 28 ml total) was used for screening (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)).

10 *T. plicata* (-)-Pinoresinol/(-)-Lарciresinol Reductase cDNA was obtained from mRNA by a reverse transcription-polymerase chain reaction (RT-PCR) strategy (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)). First-strand cDNA was synthesized from the purified mRNA previously used for the synthesis of the *T. plicata* cDNA library, described above. Purified mRNA (150 ng) was mixed with linker-primer (1.4 μ g) from ZAP-cDNA[®] synthesis kit (Stratagene), heated to 70°C for 10 min, and quickly chilled on ice. The mixture of denatured mRNA 15 template and linker-primer was then mixed with First Strand Buffer (Life Technologies), 10 mM DTT, 0.5 mM each dNTP, and 200 units of Super Script[™] II (Life Technologies) in a final volume of 20 μ l. The reaction was carried out at 42°C for 50 min and then stopped by heating (70°C, 15 min). *E. coli* RNase H (1.5 units, 20 μ l) was added to the solution and incubated at 37°C for 20 min.

20 The first-strand reaction (2 μ l) was next used as the template in 100- μ l PCR reactions (10 mM Tris-HCl, pH 9.0, 50 mM KCl, 0.1 % Triton X-100, 1.5 mM MgCl₂, 0.2 mM each dNTP, and 5 units of Tag DNA polymerase) with primer CR6-NT (5'GCACATAAGAGTATGGATAAG3')(SEQ ID No:60) (10 pmol) and primer XbaI-Poly(dT) (5'GTCTCGAGTTTTTTTTTTTTT3')(SEQ ID No:59) (10 pmol). PCR amplification was carried out in a thermocycler as described in

25

30

35

(Dinkova-Kostova, A.T., et al., *J. Biol. Chem.* 271:29473-29482 (1996)) except for the annealing temperature at 52°C. PCR products were resolved in 1.3 % agarose gels, where at least two bands possessing the expected length (about 1,200-bp) were observed. The bands were extracted from the gel. The gel-purified PCR products (56 ng) were then ligated into the pT7Blue T-vector (50 ng) and transformed into competent NovaBlue cells, according to Novagen's instructions.

The size and orientation of the inserted cDNAs were determined using the rapid boiling lysis and PCR technique, following the manufacturer's (Novagen's) instructions, with the following primer combinations: R20-mer (SEQ ID No:74) with U19-mer (SEQ ID No:75); R20-mer (SEQ ID No:74) with CR6-NT (SEQ ID No:60); U19-mer (SEQ ID No:75) with CR6-NT (SEQ ID No:60). The CR6-NT primer end of the inserted DNAs was located next to the U19-mer primer site of the T-vector. The T-vectors containing the inserted cDNAs were purified with Wizard® Plus SV Minipreps DNA Purification System. Five inserted cDNAs were completely sequenced using overlapping sequencing primers and were shown to be identical except that polyadenylation sites were different. Therefore, the longest cDNA, designated pJr-Tp1, (SEQ ID No:61) was used for detection of enzyme activity using the pBluescript expression system.

Sequence analysis - DNA and amino acid sequence analyses were performed using the Unix-based GCG Wisconsin Package (Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 (1996); Rice, P., Program Manual for the EGCG Package, Peter Rice, The Sanger Centre, Hinxton Hall, Cambridge, CB10 1Rq, England) and the ExPASy World Wide Web molecular biology server (Geneva University Hospital and University of Geneva, Geneva, Switzerland).

EXAMPLE 15

cDNA Cloning and Expression of *Thuias plicata* (+)-Pinoresinol

(+)-Lanicresinol Reductase

T. plicata (+)-Pinoresinol/(+)-Lanicresinol Reductase cDNA cloning. After pJr-Tp1 was cloned and sequenced, the full-length clone was used to screen the *T. plicata* cDNA library as described in Example 11, except that the entire pJr-Tp1 cDNA insert was used as a probe. Several positive clones were sequenced, revealing one new, unique cDNA which was called pJr-Tp2. This cDNA encodes a reductase with high sequence similarity to pJr-Tp1 (~81% similarity at the amino acid level),

but with substrate specificity properties identical to the original *Forysthia intermedia* reductase, as described below.

Enzyme Assays. Pinoresinol and lanicresinol reductase activities were assayed by monitoring the formation of [³H]lanicresinol and [³H]secoisolaricresinol as set forth in Example 8, with the following modifications. Briefly, each assay for pinoresinol reductase activity consisted of (±)-pinoresinols (5 mM in MeOH, 20 µl) and the enzyme preparation (i.e., total protein extract from *E. coli*, 210 µl). The enzymatic reaction was initiated by addition of [4R-³H]NADPH (10 mM, 6.79 kBq/nmol in distilled H₂O, 20 µl). After 3 hour incubation at 30°C with shaking, the assay mixture was extracted with EtOAc (500 µl) containing (±)-lanicresinols (20 µg) and (±)-secoisolaricresinols (20 µg) as radiochemical carriers. After centrifugation (13,800 × g, 5 min), the EtOAc solubles were removed and the extraction procedure was repeated. For each assay, the EtOAc solubles were combined with an aliquot (100 µl) removed for determination of its radioactivity using liquid scintillation counting. The remainder of the combined EtOAc solubles was evaporated to dryness *in vacuo*, reconstituted in MeOH/H₂O (30:70, 100 µl) and subjected to reversed phase and chiral column HPLC.

Lanicresinol reductase activity was assayed by monitoring the formation of (±)-[³H]secoisolaricresinol. These assays were carried out exactly as described above, except that (±)-lanicresinols (5 mM in MeOH, 20 µl) were used as substrates, with (±)-secoisolaricresinols (20 µg) added as radiochemical carriers.

Expression of pJr-Tp1 in E. coli - In order for the open reading frame (ORF) of pJr-Tp1 to be in frame with the β-galactosidase gene α-complementation particle in pBluescript SK(-), pJr-Tp1 was excised out of pT7Blue T-vector with SacI and XbaI, gel-purified, and then ligated into the expression vector digested with these same enzymes. This plasmid, pPCR-Tp1, was transformed into Novablu cells according to Novagen's instructions. The transformed cells (5-ml cultures) were grown at 37°C in LB medium (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)) supplemented with 50 µg ml⁻¹ carbenicillin with shaking (225 rpm) to mid log phase ($A_{600} = 0.5-0.7$). The cells were next collected by centrifugation (1000 × g, 10 min) and resuspended in fresh LB medium supplemented with 10 mM IPTG (isopropyl β-D-thiogalactopyranoside) and 50 µg ml⁻¹ carbenicillin to an absorbance of 0.6 (at 600 nm). The cells, allowed to grow overnight, were collected by centrifugation and resuspended in 500-700 µl of (per

5 ml culture tube) of buffer (50 mM Tris-HCl, pH 7.5, 2 mM EDTA, and 5 mM DTT). Next, the cells were lysed by sonication (5 x 45 s) and after centrifugation (17500 x g, 4°C, 10 min) the supernatant was removed and assayed for (-)-pinoresinol/(-)-larciresinol reductase activity as described above. Controls included assays of pBlue-script (SK(-)) without insert DNA (as negative control) or with pPLR-FU (cDNA of authentic *F. intermedia* (-)-pinoresinol/(-)-larciresinol reductase in frame) as stereospecific control, as well as pPLR-Tp1 with no substrate except (4R)-HNADPH.

The results showed that both (-)-larciresinol and (+)-secoisolarciresinol were radiolabelled and that no incorporation of radioactivity was found in (-)-secoisolarciresinol. However, accumulation of radiolabel into (+)-larciresinol was also observed, although at a much slower rate than that observed for (-)-larciresinol. These results indicate that pL-Tp1 can use both (-)-pinoresinol and (+)-pinoresinol as substrates, with the former being converted via (-)-larciresinol completely to (+)-secoisolarciresinol, and the latter being converted much more slowly to (-)-larciresinol, but not further to (-)-secoisolarciresinol.

Expression of pL-Tp2 in E. coli. The pL-Tp2 cDNA was found to be in frame with the β -galactosidase gene α -complementation particle in pBluescript SK(-). When evaluated for activity and substrate specificity as described above, pL-Tp2 was found to possess the same substrate specificity and product formation as the original *Forsythia intermedia* reductase (Dinkova-Kostova, A.T., et al., *J. Biol. Chem.* 271:29473-29482 (1996)) except that a small amount of (-)-larciresinol was also detected. This is interesting, because pL-Tp2 has a higher sequence similarity to pL-Tp1 than it does to the *Forstvilia* reductase.

All the above observations were confirmed using deuterolabelled substrates (\pm)-[9,9- 2 H₂, OC²H₅]pinoresinols with isolation of the corresponding lignans; each was then subjected to 'chiral' column chromatography and HPLC-mass spectral analysis to confirm these findings.

EXAMPLE 16

Cloning of Additional Pinoresinol/Larciresinol Reductases from *Thuya* *placata* and *Tsuga heterophylla*

Two additional pinoresinol/larciresinol reductases were cloned from a *Thuya placata* young stem cDNA library as described in Example 15 for the cloning of pL-Tp2. The two additional pinoresinol/larciresinol reductases were designated pL-Tp3 (SEQ ID No:65) and pL-Tp4 (SEQ ID No:67).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Lewis, Norman G
 Davin, Laurence B
 Dinkova-Kostova, Albena T
 Fujita, Masayuki
 Gang, David R
 Sarkonen, Simo

(ii) TITLE OF INVENTION: Recombinant Pinoresinol/Lariciresinol Reductases, Recombinant Dirigent Proteins and Methods of Use

(iii) NUMBER OF SEQUENCES: 76

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Christensen, O'Connor, Johnson & Kindness
 (B) STREET: 1420 Fifth Avenue, Suite 2800
 (C) CITY: Seattle
 (D) STATE: Washington
 (E) COUNTRY: USA
 (F) ZIP: WA 98101-2347

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: Patentin Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
 (B) FILING DATE:
 (C) CLASSIFICATION:

(vii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Shelton, Dennis K
 (B) REGISTRATION NUMBER: 26,997
 (C) REFERENCE/DOCKET NUMBER: WSUR111351

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 206 682 8100
 (B) TELEFAX: 206 224 0779

(x) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment
 (vi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

(vii) FRAGMENT TYPE: N-terminal
 (viii) ORIGINAL SOURCE:
 (A) ORGANISM: Forsythia intermedia dirigent protein N-terminal sequence

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Lys Pro Arg Pro Xaa Arg Xaa Xaa Lys Glu Leu Val Phe Tyr Phe Xaa
 1 5 10 15
 Asp Ile Leu Phe Lys Gly Xaa Asn Tyr Asn Xaa Asa
 20 25

(x) INFORMATION FOR SEQ ID NO:2:

(xi) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: not relevant

(xii) MOLECULE TYPE: peptide

(xiii) HYPOTHETICAL: NO
 (xiv) ANTI-SENSE: NO

(xv) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment

(xvi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Thr Ala Met Ala Val Pro Phe Asn Tyr Gly Asp Leu Val Val Phe Asp
 1 5 10 15
 Asp Pro Ile Thr Leu Asp Asn Asn

(xvii) INFORMATION FOR SEQ ID NO:3:

(xviii) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: not relevant

(xix) MOLECULE TYPE: peptide

(xx) HYPOTHETICAL: NO
 (xxi) ANTI-SENSE: NO

(xxii) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment

(xxiii) SEQUENCE DESCRIPTION: SEQ ID NO:3:

-69-

Tyr Val Gly Thr Leu Asn Phe Ala Gly Ala Asp Pro Leu Leu Xaa Lys
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:4:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: *Forsythia intermedia* dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asp Ile Ser Val Ile Gly Thr Gly Asp Phe Phe Met Ala Arg
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:5:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: *Forsythia intermedia* dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Gly Val Ala Thr Leu Met Thr Asp Ala Phe Glu Gly Asp Xaa Tyr
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:6:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide

Tyr Val Gly Thr Leu Asn Phe Ala Gly Ala Asp Pro Leu Leu Xaa Lys
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: *Forsythia intermedia* dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ala Gin Gly Met Tyr Phe Tyr Asp Gln Lys
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) SEQUENCE DESCRIPTION: "PCR primer PSINT1"

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) SEQUENCE DESCRIPTION: "PCR primer PSINT2"

-71-

PCT/US97/20391 -

-72-

(A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i.i) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: "PCR primer PS1R"

(ii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TARTTRANG GRAGNCGCAT

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i.i) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: "PCR primer PS12R"

(ii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTNATNGGRT CRTCRAANAC

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: "PCR primer PS1R"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CCATRARAAT RTCNCNCNT

(C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (iii) MOLECULE TYPE: cDNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: *Forsythia intermedia* clone psd-fil
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 26..583
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
 ATTTGGCAG GAGGTTAAC CAAAC ATG GTT TCT AAA ACA CAA ATT GTC GCT
 52
 Met Val Ser Lys Thr Gln Ile Val Ala
 CTT TCC CTT TGC TTC CTC ACT TCC ACC TCT TCC GCC ACC TAC GGC CGC
 Leu Phe Leu Cys Phe Leu Thr Ser Ser Ala Thr Tyr Gly Arg
 10 15 20 25
 AAG CCA CGC CCT CGG CGG CCC TGC AAA GAA TTG GTG TTC TAT TTC CAC
 Lys Pro Arg Pro Arg Pro Cys Lys Glu Leu Val Phe Tyr Phe His
 30 35 40
 GAC GTA CTT TTC AAA GGA AAT ATT TAC CAC AAT GCC ACT TCC CGC ATA
 Asp Val Leu Phe Lys Glu Asn Asn Tyr His Asn Ala Thr Gln Ile
 45 50 55
 GTC GGG TCC CCC CAA TGG GGC AAC AAG ACT GCC ATG GCC GTG CCA TTC
 Val Gly Ser Pro Glu Trp Gly Asn Lys Thr Ala Met Ala Val Pro Phe
 60 65 70
 AAT TAT GGT GAC CTA GTT GTG TTC GAC GAT CCC ATT ACC TTA GAC AAC
 Asn Tyr Gly Asp Leu Val Phe Asp Asp Pro Ile Thr Leu Asp Asn
 75 80 85
 AAT CTG CAT TCA CCC CCA GTG GGT CGG CGG CAA GGG ATG TAC TTC TAT
 Asn Leu His Ser Pro Val Glu Arg Ala Gln Gly Met Tyr Phe Tyr
 90 95 100 105
 GAT CAA AAA AAT ACA TAC ATT GCT TGG CTA GGG TTC TCA TTT TTG TTC
 Asp Gln Lys Asn Thr Tyr Asn Ala Trp Leu Gly Phe Ser Phe Ile
 110 115 120
 AAT TCA ACT AAG TAT GTT GGA ACC TTG AAC TTT GCT GGG GCT GAT CCA
 Asn Ser Thr Lys Tyr Val Glu Thr Leu Asn Phe Ala Gly Ala Asp Pro
 125 130 135
 TTG TTG AAC AAG ACT AGA GAC ATA TCA GTC ATT GGT GGA ACT GGT GAC
 Leu Leu Asn Lys Thr Arg Asp Ile Ser Val Ile Gly Glu Thr Gly Asp
 140 145 150
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 901 base pairs
 (B) TYPE: nucleic acid

(2) INFORMATION FOR SEQ ID NO:12:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 901 base pairs
 (B) TYPE: nucleic acid

19

(C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (iii) MOLECULE TYPE: cDNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: *Forsythia intermedia* clone psd-fil
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 26..583
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
 ATTTGGCAG GAGGTTAAC CAAAC ATG GTT TCT AAA ACA CAA ATT GTC GCT
 52
 Met Val Ser Lys Thr Gln Ile Val Ala
 CTT TCC CTT TGC TTC CTC ACT TCC ACC TCT TCC GCC ACC TAC GGC CGC
 Leu Phe Leu Cys Phe Leu Thr Ser Ser Ala Thr Tyr Gly Arg
 10 15 20 25
 AAG CCA CGC CCT CGG CGG CCC TGC AAA GAA TTG GTG TTC TAT TTC CAC
 Lys Pro Arg Pro Arg Pro Cys Lys Glu Leu Val Phe Tyr Phe His
 30 35 40
 GAC GTA CTT TTC AAA GGA AAT ATT TAC CAC AAT GCC ACT TCC CGC ATA
 Asp Val Leu Phe Lys Glu Asn Asn Tyr His Asn Ala Thr Gln Ile
 45 50 55
 GTC GGG TCC CCC CAA TGG GGC AAC AAG ACT GCC ATG GCC GTG CCA TTC
 Val Gly Ser Pro Glu Trp Gly Asn Lys Thr Ala Met Ala Val Pro Phe
 60 65 70
 AAT TAT GGT GAC CTA GTT GTG TTC GAC GAT CCC ATT ACC TTA GAC AAC
 Asn Tyr Gly Asp Leu Val Phe Asp Asp Pro Ile Thr Leu Asp Asn
 75 80 85
 AAT CTG CAT TCA CCC CCA GTG GGT CGG CGG CAA GGG ATG TAC TTC TAT
 Asn Leu His Ser Pro Val Glu Arg Ala Gln Gly Met Tyr Phe Tyr
 90 95 100 105
 GAT CAA AAA AAT ACA TAC ATT GCT TGG CTA GGG TTC TCA TTT TTG TTC
 Asp Gln Lys Asn Thr Tyr Asn Ala Trp Leu Gly Phe Ser Phe Ile
 110 115 120
 AAT TCA ACT AAG TAT GTT GGA ACC TTG AAC TTT GCT GGG GCT GAT CCA
 Asn Ser Thr Lys Tyr Val Glu Thr Leu Asn Phe Ala Gly Ala Asp Pro
 125 130 135
 TTG TTG AAC AAG ACT AGA GAC ATA TCA GTC ATT GGT GGA ACT GGT GAC
 Leu Leu Asn Lys Thr Arg Asp Ile Ser Val Ile Gly Glu Thr Gly Asp
 140 145 150
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 901 base pairs
 (B) TYPE: nucleic acid

-73-

TTT TTC ARG GCG AGA GGG GTC PCT TTG ARG ACC GAT GCC TTT GAA
Phe Met Ala Arg Gly Val Ala Thr Asp Ala Phe Glu 532

GGY GAT GAG TAT TCC CGC CTC GTC GAT GTC GAT ATT ATT TGT TAT GAA TGT
Gly Asp Val Tyr Phe Arg Leu Arg Val Asp Ile Asn Leu Tyr Glu Cys 580

TGG TAAACATT AGCCGATAT ATATATATAT ATGGCTTATAC ATATTCATA
TTP 633

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 186 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: protein: Forsythia intermedia PSD-F11 protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Val Ser Lys Thr Gln Ile Val Ala Leu Phe Leu Cys Phe Leu Thr 1
5 10 15
Ser Thr Ser Ala Thr Tyr Gly Arg Pro Arg Arg Pro 20 25
Cys Lys Glu Leu Val Phe His Asp Val Ile Phe Lys Gly Asn 35 40 45
Asn Tyr His Asn Ala Thr Ser Ala Ile Val Gly Ser Pro Gln Trp Gly 50 55 60
Asn Lys Thr Ala Met Ala Val Pro Phe Asn Tyr Gly Asp Leu Val Val 65 70 75 80
Phe Asp Asp Pro Ile Thr Leu Asp Asn His Ser Pro Pro Val 85 90 95
Gly Arg Ala Gln Gly Met Tyr Phe Tyr Asp Gln Lys Asn Thr Tyr Asn 100 105 110
Ala Trp Leu Gly Phe Ser Phe Leu Phe Asn Ser Thr Lys Tyr Val Gly 115 120 125
Thr Leu Asn Phe Ala Gly Ala Asp Pro Leu Leu Asn Lys Thr Arg Asp 130 135 140

Ile Ser Val Ile Gly Gly Asp Phe Phe Met Ala Arg Gly Val
145 150 155 160
Ala Thr Leu Met Thr Asp Ala Phe Glu Gly Asp Val Tyr Phe Arg Leu 165 170 175
Arg Val Asp Ile Asn Leu Tyr Glu Cys Trp 180 185 185

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 858 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia cDNA PSD-F12

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 19..573

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
AATTCGGCAC GAGGAAATG GCA GCT AAA ACA CAA ACC ACA GCC CTTT TTC
Met Ala Ala Iys Thr Gln Thr Thr Ala Leu Phe 51
CTC TGC CTC CTC ATC TGC ATC TCC CCC GTG TAC GGC CAC AAA ACC AGG
Leu Cys Leu Ile Cys Ile Ser Ala Val Tyr Gly His Lys Thr Arg 99
200
TCT CGA CGC CCC TGT AAA GAG CTC GTC GTC TCC CAC GAC ATC CTC
Ser Arg Arg Pro Cys Lys Glu Leu Val Phe His Asp Ile Leu 147
215 220
TAC CTA GGA TAC RAT AGA AAC ATC GCC ACC GCT GTC ATA GTC GCC TCT
Tyr Leu Gly Tyr Asn Arg Asn Ala Thr Ala Val Ile Val Ala Ser 195
230 235 240 245
CCT CRA TCG GGA AAC AAC RCT GCC ATG GCT AAA CCT TTC AAT TTT GGT
Pro Gln Trp Gly Asn Lys Thr Ala Met Ala Lys Pro Phe Gly 243
250 255 260
GAT TGG GTC GTC TTT GAT GTC CCC ATT ACC TTA GTC AAC AAC CTG CAT
Asp Leu Val Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His 291
265 270 275
TCT CCT CCG GTC GGC CGG GCT CAG GGA ACT TAT TTC TAC GAT CAA TGG
Ser Pro Pro Val Gly Arg Ala Glu Gly Thr Tyr Phe Tyr Asp Glu Trp 339
280 285 290

-74-

-75-

AGT ATT TAT GGT GCA TGG CTT GGA TTT TCA TTT TGT TTC AAT TCT ACT
Ser Ile Tyr Gly Ala Trp Leu Gly Phe Ser Phe Ile Leu Phe Asn Ser Thr
295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370

GAT TAT GTT GGA ACT CTA AAT TTT GCT GGA GCT GAT CCA TTT ATT AAC
Asp Tyr Val Gly Thr Ile Asn Phe Ala Gly Ala Asp Pro Leu Ile Asn
310 315 320 325 330 335 340 345 350 355 360 365 370

AAA ACT AGG GAC ATT TCA GTC ATT GGA GCA ACT GCT GAT TTT TTC ATT
Lys Thr Arg Asp Ile Ser Val Ile Gly Thr Gly Asp Phe Met
330 335 340 345 350 355 360 365 370

GCT AGA GGG GTC GCC ACT GTG TCG ACC GAT GCT TTT GAA GGG GAT GTT
Ala Arg Gly Val Ala Thr Val Ser Thr Asp Ala Phe Glu Gly Asp Val
340 345 350 355 360 365 370

TAT TTC AGG CTT CGT GTT GAT ATT AGG TTG TAT GAG TGT TGG
Tyr Phe Arg Leu Arg Val Asp Ile Arg Leu Tyr Glu Cys Trp
360 365 370

TAATTTTACCTTATTTTCCATTTTTCTGAATTTTCTGAATTTGACGCGATTGAGACTAATTATGCTT
633.

CTGTTATCTTGTTTGTATCTATTGTGGCGATTTTATCATTAGTGATTTGTTGGTC
693

ATATTTTATCTGTAAAAAAATTATGGTCAAAAGGCCAACTACACACCTAGGGACT
753

TTTTTCCGTTTACGGGAAAAAABAAAGATGGTCTTACGAGTGTAACTACGTTTCATTICATIC
813

AAAAATTGGCTTTCAATCATCTTCCTCATAAAARAAAAAAARAAAA
858

(2) INFORMATION FOR SEQ ID NO:15:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 185 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: Forysthia intermedia dirigent protein PSD-F12
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Ala Ala Lys Thr Gln Thr Thr Ala Leu Phe Leu Ile
1 5 10 15

Cys Ile Ser Ala Val Tyr Gly His Lys Thr Arg Ser Arg Pro Cys
20 25 30

Lys Glu Leu Val Phe Phe His Asp Ile Leu Tyr Leu Gly Tyr Asn
35 40 45 50 55 60

Arg Asn Asn Ala Thr Ala Val Ile Val Ala Ser Pro Gln Trp Gly Asn
65 70 75 80

Lys Thr Ala Met Ala Lys Pro Phe Asn Phe Gly Asp Leu Val Phe
85 90 95

Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Val Gly
Tyr Asn Gly Ser Asn Ala Lys Asn Ala Thr Ser Thr Leu Val Gly Ala
240 250 255 260

-76-

Arg Ala Gln Gly Thr Tyr Phe Tyr Asp Gln Trp Ser Ile Tyr Gly Ala
100 105 110

Trp Leu Gly Phe Ser Phe Leu Phe Asn Ser Thr Asp Tyr Val Gly Thr
115 120 125

Leu Asn Phe Ala Gly Ala Asp Pro Leu Ile Asn Lys Thr Arg Asp Ile
130 135 140

Ser Val Ile Gly Thr Gly Asp Phe Phe Met Ala Arg Gly Val Ala
145 150 155 160

Thr Val Ser Thr Asp Ala Phe Glu Gly Asp Val Tyr Phe Arg Leu Arg
165 170 175

Val Asp Ile Arg Leu Tyr Glu Cys Trp
180 185

(2) INFORMATION FOR SEQ ID NO:16:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 948 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: Tsuga heterophylla dirigent protein cDNA PSD-Th1
(ii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 104..688

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGGCACCCCTCCTGTGTTATCTGAGCCCTTCCTCTCTACTCTCTCTGTTAATCTCTTGTAT
60

CCCGATATCTTCTCTATATCACTTAGTCCTATAAGATGTCTATGAACTGAAATGAAATCAG
115

AAT CGT AAT AGA GCT GTG CAC TTG TGT TTT CTA TGT CTT CTA CTC TCC
Asn Arg Asn Arg Ala Val His Leu Cys Phe Leu Trp Leu Leu Ser
190 195 200 205

TCT GTG TTG CAA AGA ACT GAT GAT GGG AAA AGC TTG AAG AAC CAC GCA
Ser Val Leu Leu Gln Thr Ser Asp Gly Lys Ser Trp Lys His Arg
210 215 220

CTC CGA AAG CCT TGT AGG AAT CTG GTG TTG TAT TTC CAT GAT GTC ATC
Leu Arg Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp Val Ile
225 230 235

TAC AAT GGC AGC AAC GGC AAG AAC GCT ACA TCC ACA CTT GTG GGT CCT
Tyr Asn Gly Ser Asn Ala Lys Asn Ala Thr Ser Pro Val Gly
240 250 255 260

-77-

-78-

CCC CAC GGG TCT AAC CTC ACA CTT CTC GCT GGA AAA GAC AAC CAC TTT
Pro His .Gly Ser Asn Leu Thr Leu Ala Gly Lys Asn His Phe
255 260

GGA GAT CTG GCG GTG TTT GAC GAT CCC ATC ACT CTT GAC AAC AAT TTC
Gly Asp Leu Ala Val The Asp Asp Pro Ile Thr Leu Asp Asn Phe
270 275 280

CAC TCT CCT CGG GRG GGC AGA GCT CAG GGA TTC TAC TTT TAT GAC ATG
His Ser Pro Val Gly Arg Ala Gly Phe Tyr Asp Mat
290 295 300

AAG AAC ACC TMC AGC TCC TGG CTT GGA TTC ACG TTT GTC CTC AAC TCT
Lys Asn Thr Phe Ser Ser Trp Leu Gly Phe Thr Phe Val Leu Asn Ser
305 310

ACA GAT TAC ARA GGC ACC ATC AGC TTC TCT GGA GCC GAT CCA ATC CTT
Thr Asp Tyr Lys Gly Thr Ile Thr Phe Ser Gly Ala Asp Pro Ile Leu
320 325

ATC AAA TAC AGA GAT ATA TCA GTG GGA ACT GGA GAT TGC ATA
Thr Lys Tyr Arg Asp Ile Ser Val Gly Gly Asp Phe Ile
335 340

ATG GCA AGA GGA ATC GCC ACA ATC TCC RCC GAT TCG TAT GAA GGC GAC
Met Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ala Tyr Glu Gly Asp
350 355

GTT TAC TCC CGT CTC TGC GTG AAT ATC ACA CTC TAT GAG TGC TAC
Val Tyr Phe Arg Leu Cys Val Asn Ile Thr Leu Tyr Glu Cys Tyr
370 375

TGAGTCCTATT AGGTCTTATT TCTCCCTCGA CTATCCATT ATTATGTCAT TTGTAGTGA
CTAGTGTATT CTTGTCGAG AGAATGCACT
808

CTTGTGAGCC CGAAATGATT ATTGTGATT TGTGCAAGGC CATATCTPAT ATTGTGCAAGG
GAAATGCGA ATTCTCTATTG CGTCAGGCA CTTTATTAATAATAAGG ATATATTGGT
TATATATAAA AAATATAA
928

948

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 195 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: *Tsuga heterophylla* dirigent protein PSD-Th1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Ala Ile Lys Asn Arg Ala Val His Leu Cys Phe Ile Trp
1 5 10 15

Leu Leu Ser Ser Val Leu Gln Thr Ser Asp Gly Lys Ser Trp
20 25 30

355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080 1085 1090 1095 1100 1105 1110 1115 1120 1125 1130 1135 1140 1145 1150 1155 1160 1165 1170 1175 1180 1185 1190 1195 1200 1205 1210 1215 1220 1225 1230 1235 1240 1245 1250 1255 1260 1265 1270 1275 1280 1285 1290 1295 1300 1305 1310 1315 1320 1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 1395 1400 1405 1410 1415 1420 1425 1430 1435 1440 1445 1450 1455 1460 1465 1470 1475 1480 1485 1490 1495 1500 1505 1510 1515 1520 1525 1530 1535 1540 1545 1550 1555 1560 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 1615 1620 1625 1630 1635 1640 1645 1650 1655 1660 1665 1670 1675 1680 1685 1690 1695 1700 1705 1710 1715 1720 1725 1730 1735 1740 1745 1750 1755 1760 1765 1770 1775 1780 1785 1790 1795 1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865 1870 1875 1880 1885 1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050 2055 2060 2065 2070 2075 2080 2085 2090 2095 2100 2105 2110 2115 2120 2125 2130 2135 2140 2145 2150 2155 2160 2165 2170 2175 2180 2185 2190 2195 2200 2205 2210 2215 2220 2225 2230 2235 2240 2245 2250 2255 2260 2265 2270 2275 2280 2285 2290 2295 2300 2305 2310 2315 2320 2325 2330 2335 2340 2345 2350 2355 2360 2365 2370 2375 2380 2385 2390 2395 2400 2405 2410 2415 2420 2425 2430 2435 2440 2445 2450 2455 2460 2465 2470 2475 2480 2485 2490 2495 2500 2505 2510 2515 2520 2525 2530 2535 2540 2545 2550 2555 2560 2565 2570 2575 2580 2585 2590 2595 2600 2605 2610 2615 2620 2625 2630 2635 2640 2645 2650 2655 2660 2665 2670 2675 2680 2685 2690 2695 2700 2705 2710 2715 2720 2725 2730 2735 2740 2745 2750 2755 2760 2765 2770 2775 2780 2785 2790 2795 2800 2805 2810 2815 2820 2825 2830 2835 2840 2845 2850 2855 2860 2865 2870 2875 2880 2885 2890 2895 2900 2905 2910 2915 2920 2925 2930 2935 2940 2945 2950 2955 2960 2965 2970 2975 2980 2985 2990 2995 3000 3005 3010 3015 3020 3025 3030 3035 3040 3045 3050 3055 3060 3065 3070 3075 3080 3085 3090 3095 3100 3105 3110 3115 3120 3125 3130 3135 3140 3145 3150 3155 3160 3165 3170 3175 3180 3185 3190 3195 3200 3205 3210 3215 3220 3225 3230 3235 3240 3245 3250 3255 3260 3265 3270 3275 3280 3285 3290 3295 3300 3305 3310 3315 3320 3325 3330 3335 3340 3345 3350 3355 3360 3365 3370 3375 3380 3385 3390 3395 3400 3405 3410 3415 3420 3425 3430 3435 3440 3445 3450 3455 3460 3465 3470 3475 3480 3485 3490 3495 3500 3505 3510 3515 3520 3525 3530 3535 3540 3545 3550 3555 3560 3565 3570 3575 3580 3585 3590 3595 3600 3605 3610 3615 3620 3625 3630 3635 3640 3645 3650 3655 3660 3665 3670 3675 3680 3685 3690 3695 3700 3705 3710 3715 3720 3725 3730 3735 3740 3745 3750 3755 3760 3765 3770 3775 3780 3785 3790 3795 3800 3805 3810 3815 3820 3825 3830 3835 3840 3845 3850 3855 3860 3865 3870 3875 3880 3885 3890 3895 3900 3905 3910 3915 3920 3925 3930 3935 3940 3945 3950 3955 3960 3965 3970 3975 3980 3985 3990 3995 4000 4005 4010 4015 4020 4025 4030 4035 4040 4045 4050 4055 4060 4065 4070 4075 4080 4085 4090 4095 4100 4105 4110 4115 4120 4125 4130 4135 4140 4145 4150 4155 4160 4165 4170 4175 4180 4185 4190 4195 4200 4205 4210 4215 4220 4225 4230 4235 4240 4245 4250 4255 4260 4265 4270 4275 4280 4285 4290 4295 4300 4305 4310 4315 4320 4325 4330 4335 4340 4345 4350 4355 4360 4365 4370 4375 4380 4385 4390 4395 4400 4405 4410 4415 4420 4425 4430 4435 4440 4445 4450 4455 4460 4465 4470 4475 4480 4485 4490 4495 4500 4505 4510 4515 4520 4525 4530 4535 4540 4545 4550 4555 4560 4565 4570 4575 4580 4585 4590 4595 4600 4605 4610 4615 4620 4625 4630 4635 4640 4645 4650 4655 4660 4665 4670 4675 4680 4685 4690 4695 4700 4705 4710 4715 4720 4725 4730 4735 4740 4745 4750 4755 4760 4765 4770 4775 4780 4785 4790 4795 4800 4805 4810 4815 4820 4825 4830 4835 4840 4845 4850 4855 4860 4865 4870 4875 4880 4885 4890 4895 4900 4905 4910 4915 4920 4925 4930 4935 4940 4945 4950 4955 4960 4965 4970 4975 4980 4985 4990 4995 5000 5005 5010 5015 5020 5025 5030 5035 5040 5045 5050 5055 5060 5065 5070 5075 5080 5085 5090 5095 5100 5105 5110 5115 5120 5125 5130 5135 5140 5145 5150 5155 5160 5165 5170 5175 5180 5185 5190 5195 5200 5205 5210 5215 5220 5225 5230 5235 5240 5245 5250 5255 5260 5265 5270 5275 5280 5285 5290 5295 5300 5305 5310 5315 5320 5325 5330 5335 5340 5345 5350 5355 5360 5365 5370 5375 5380 5385 5390 5395 5400 5405 5410 5415 5420 5425 5430 5435 5440 5445 5450 5455 5460 5465 5470 5475 5480 5485 5490 5495 5500 5505 5510 5515 5520 5525 5530 5535 5540 5545 5550 5555 5560 5565 5570 5575 5580 5585 5590 5595 5600 5605 5610 5615 5620 5625 5630 5635 5640 5645 5650 5655 5660 5665 5670 5675 5680 5685 5690 5695 5700 5705 5710 5715 5720 5725 5730 5735 5740 5745 5750 5755 5760 5765 5770 5775 5780 5785 5790 5795 5800 5805 5810 5815 5820 5825 5830 5835 5840 5845 5850 5855 5860 5865 5870 5875 5880 5885 5890 5895 5900 5905 5910 5915 5920 5925 5930 5935 5940 5945 5950 5955 5960 5965 5970 5975 5980 5985 5990 5995 6000 6005 6010 6015 6020 6025 6030 6035 6040 6045 6050 6055 6060 6065 6070 6075 6080 6085 6090 6095 6100 6105 6110 6115 6120 6125 6130 6135 6140 6145 6150 6155 6160 6165 6170 6175 6180 6185 6190 6195 6200 6205 6210 6215 6220 6225 6230 6235 6240 6245 6250 6255 6260 6265 6270 6275 6280 6285 6290 6295 6300 6305 6310 6315 6320 6325 6330 6335 6340 6345 6350 6355 6360 6365 6370 6375 6380 6385 6390 6395 6400 6405 6410 6415 6420 6425 6430 6435 6440 6445 6450 6455 6460 6465 6470 6475 6480 6485 6490 6495 6500 6505 6510 6515 6520 6525 6530 6535 6540 6545 6550 6555 6560 6565 6570 6575 6580 6585 6590 6595 6600 6605 6610 6615 6620 6625 6630 6635 6640 6645 6650 6655 6660 6665 6670 6675 6680 6685 6690 6695 6700 6705 6710 6715 6720 6725 6730 6735 6740 6745 6750 6755 6760 6765 6770 6775 6780 6785 6790 6795 6800 6805 6810 6815 6820 6825 6830 6835 6840 6845 6850 6855 6860 6865 6870 6875 6880 6885 6890 6895 6900 6905 6910 6915 6920 6925 6930 6935 6940 6945 6950 6955 6960 6965 6970 6975 6980 6985 6990 6995 7000 7005 7010 7015 7020 7025 7030 7035 7040 7045 7050 7055 7060 7065 7070 7075 7080 7085 7090 7095 7100 7105 7110 7115 7120 7125 7130 7135 7140 7145 7150 7155 7160 7165 7170 7175 7180 7185 7190 7195 7200 7205 7210 7215 7220 7225 7230 7235 7240 7245 7250 7255 7260 7265 7270 7275 7280 7285 7290 7295 7300 7305 7310 7315 7320 7325 7330 7335 7340 7345 7350 7355 7360 7365 7370 7375 7380 7385 7390 7395 7400 7405 7410 7415 7420 7425 7430 7435 7440 7445 7450 7455 7460 7465 7470 7475 7480 7485 7490 7495 7500 7505 7510 7515 7520 7525 7530 7535 7540 7545 7550 7555 7560 7565 7570 7575 7580 7585 7590 7595 7600 7605 7610 7615 7620 7625 7630 7635 7640 7645 7650 7655 7660 7665 7670 7675 7680 7685 7690 7695 7700 7705 7710 7715 7720 7725 7730 7735 7740 7745 7750 7755 7760 7765 7770 7775 7780 7785 7790 7795 7800 7805 7810 7815 7820 7825 7830 7835 7840 7845 7850 7855 7860 7865 7870 7875 7880 7885 7890 7895 7900 7905 7910 7915 7920 7925 7930 7935 7940 7945 7950 7955 7960 7965 7970 7975 7980 7985 7990 7995 8000 8005 8010 8015 8020 8025 8030 8035 8040 8045 8050 8055 8060 8065 8070 8075 8080 8085 8090 8095 8100 8105 8110 8115 8120 8125 8130 8135 8140 8145 8150 8155 8160 8165 8170 8175 8180 8185 8190 8195 8200 8205 8210 8215 8220 8225 8230 8235 8240 8245 8250 8255 8260 8265 8270 8275 8280 8285 8290 8295 8300 8305 8310 8315 8320 8325 8330 8335 8340 8345 8350 8355 8360 8365 8370 8375 8380 8385 8390 8395 8400 8405 8410 8415 8420 8425 8430 8435 8440 8445 8450 8455 8460 8465 8470 8475 8480 8485 8490 8495 8500 8505 8510 8515 8520 8525 8530 8535 8540 8545 8550 8555 8560 8565 8570 8575 8580 8585 8590 8595 8600 8605 8610 8615 8620 8625 8630 8635 8640 8645 8650 8655 8660 8665 8670 8675 8680 8685 8690 8695 8700 8705 8710 8715 8720 8725 8730 8735 8740 8745 8750 8755 8760 8765 8770 8775 8780 8785 8790 8795 8800 8805 8810 8815 8820 8825 8830 8835 8840 8845 8850 8855 8860 8865 8870 8875 8880 8885 8890 8895 8900 8905 8910 8915 8920 8925 8930 8935 8940 8945 8950 8955 8960 8965 8970 8975 8980 8985 8990 8995 9000 9005 9010 9015 9020 9025 9030 9035 9040 9045 9050 9055 9060 9065 9070 9075 9080 9085 9090 9095 9100 9105 9110 9115 9120 9125 9130 9135 9140 9145 9150 9155 9160 9165 9170 9175 9180 9185 9190 9195 9200 9205 9210 9215 9220 9225 9230 9235 9240 9245 9250 9255 9260 9265 9270 9275 9280 9285 9290 9295 9300 9305 9310 9315 9320 9325 9330 9335 9340 9345 9350 9355 9360 9365 9370 9375 9380 9385 9390 9395 9400 9405 9410 9415 9420 9425 9430 9435 9440 9445 9450 9455 9460 9465 9470 9475 9480 9485 9490 9495 9500 9505 9510 9515 9520 9525 9530 9535 9540 9545 9550 9555 9560 9565 9570 9575 9580 9585 9590 9595 9600 9605 9610 9615 9620 9625 9630 9635 9640 9645 9650 9655 9660 9665 9670 9675 9680 9685 9690 9695 9700 9705 9710 9715 9720 9725 9730 9735 9740 9745 9750 9755 9760 9765 9770 9775 9780 9785 9790 9795 9800 9805 9810 9815 9820 9825 9830 9835 9840 9845 9850 9855 9860 9865 9870 9875 9880 9885 9890 9895 9900 9905 9910 9915 9920 9925 9930 9935 9940 9945 9950 9955 9960 9965 9970 9975 9980 9985 9990 9995 9999

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 849 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: *Tsuga heterophylla* dirigent protein PSD-Th1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Ala Ile Lys Asn Arg Ala Val His Leu Cys Phe Ile Trp
1 5 10 15

Leu Leu Ser Ser Val Leu Gln Thr Ser Asp Gly Lys Ser Trp
20 25 30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 195 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: *Tsuga heterophylla* dirigent protein PSD-Th1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GTTCTGTC AAATCTTAAT TAGCCTTCAGA TTCAATCCAG GATCCCACTC TTCTTCCTTC
60

AAGATGGCA ATG ATC AAG AGT AAT AGG GCT GTG TTG CGT TGC TTT
109

Met Ala Ile Lys Ser Asp Gly Lys Ser Trp
200 205

GTA TGG CTT CTG TTG CAA AGT GGT TTT GTC CTC CCA CAG
Val Trp Leu Leu Leu Leu Gln Ser Gly Phe Val Phe Pro Leu Pro Gln
210 215 220

CCT TGT AGG AAT CTG TTG TAT TTC CAC GAT GTC CTC TAC ATT GGC
Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp Val Leu Tyr Asn Gly
225 230 235 240

TTC AAC GGC CAC AAC GCT ACA CTT GTC GGT GCT GCA CAG GGG
Phe Asn Ala His Asn Ala Thr Ser Thr Leu Val Gly Ala Pro Gln Gly
245 250 255 260

GCT AAC CTC ACA CTT CTC GCT GGA AAA GAC AAC CAC TTT GGA GAT CTG
Ala Asn Leu Thr Leu Leu Ala Gly Lys Asp Asn His Phe Gly Asp Leu
265 270 275 280

GCG GTG TCC GAC GAT CCC ATC ACT CTT GAC AAC ATC TTC CAG TCT CCT
Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Phe Gln Ser Pro
275 280 285

CCG GTG GGC AGA GCT CAG GGA TTC TAC TTT TAT GAC ATG AAG AAC ACC
Pro Val Gly Arg Ala Gln Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asn Thr
290 295 300

ATC AGC TCC TGG CTT GGA TTC ACG TTT GTC CTC ABC TCT AGA GAT TAC
Phe Ser Ser Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp Tyr
305 310 315 320

AAA GGC ACC ATC AGC TTC TCT GGA GCC GAT CCA ATC CTT ACT AAA TAC
Lys Gly Thr Ile Thr Phe Ser Gly Ala Asp Pro Ile Leu Thr Lys Tyr
325 330 335 340

AGA GAT ATA TCA GTG GGA ACT GGA GAT TTC ATA ATG GCA AGA
Arg Asp Ile Ser Val Val Gly Gly Tyr Gly Asp Phe Ile Met Ala Arg
340 345 350

GGA ATC GCC ACA ATC TCC ACC GAT GGC TAT GAA GGA GAT GTC TAC
Gly Ile Ala Thr Ile Ser Thr Asp Ala Tyr Glu Gly Asp Val Tyr Phe
355 360 365

CGT CTC CGC GTC ATC ACA CTC TAT GAA TGC TAC TGTATTTAT
Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
370 375 380

NAGTAGCTAC TGTGTTCTCGT CTGGCTCGC CATTGCGATG CTCTTTAA CATTAGTGCT
TTCCATATAT TGTGTAGCC TCTCTATATA ACCCGATAA ATATTCCTTC TTGTATTA
GCAGGTTTCA ATGCTATGTA TTAGTATTTAT TTATTTATG GATTTATAC AGTCATATA
AATATTTCT CAGCTAAAAA AAAAAAAA AAAA

(iii) MOLECULE TYPE: *Tsuga heterophylla* dirigent protein translated
from PSD-Th2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met Ala Ile Lys Ser Asn Arg Ala Val Arg Phe Cys Phe Val Trp Leu
1 5 10 15

Leu Leu Gln Ser Gly Phe Val Phe Pro Leu Pro Gln Pro Cys Arg
20 25 30

Asn Leu Val Leu Tyr Phe His Asp Val Leu Tyr Asn Gly Phe Asn Ala
35 40 45

His Asn Ala Thr Ser Thr Leu Val Gly Ala Pro Gln Gly Ala Asn Leu
50 55 60

Thr Leu Leu Ala Gly Lys Asp Asn His Phe GLY Asp Leu Ala Val Phe
65 70 75 80

Asp Asp Pro Ile Thr Leu Asp Asn Phe Gln Ser Pro Pro Val Gly
85 90 95

Arg Ala Gln Gly Phe Tyr Asp Met Lys Asn Thr Phe Ser Ser
100 105 110

Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp Tyr Lys Gly Thr
115 120 125

Ile Thr Phe Ser Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg Asp Ile
130 135 140

Ser Val Val Gly Gly Thr Gly Asp Phe Ile Met Ala Arg Gly Ile Ala
145 150 155 160

Thr Ile Ser Thr Asp Ala Tyr Glu Gly Asp Val Tyr Phe Arg Leu Arg
165 170 175

Val Asn Ile Thr Leu Tyr Glu Cys Tyr
180 185

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 873 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: *Thuja plicata* dirigent protein PSD-Tpl cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 25..591

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 185 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

-81-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGTGGACGAG GGATTCGAAG AGAT AGT AGT AGA ATA GCA TTT CAT TTG TGC
Phe Met Gly Leu Leu Ser Ser Thr Val Leu Arg Asn Val Asp Gly 51
195 200 205 210
CAT GCA TGG AAG AGG CAA CCTT CCA ATG CCA TGT AAG AAT TTG GTG CTC
His Ala-Tyrp Lys Arg Gln Leu Pro Met Pro Cys Lys Asn Leu Val Ieu 147
215 220 225
TAC TTT CAT GAT GAT ATA CTC TAC AAT GGC AAA AAC ATT CAC AAT GCA ACT
Tyr Phe His Asp Ile Leu Tyr Asn Gly Lys Asn Ile His Asn Ala Thr 195
230 235 240
GCT GCG CTC GTT GCA GCT CCT CGT GGC AAT CTC ACT ACT TIC GCT
Ala Ala Leu Val Ala Ala Pro Ala Tyrp Gly Asn Leu Thr Thr Phe Ala 243
245 250 255
GAA CCT TTC ARG TTT GGA GAT GTC GTC GTC GAT CCC ATT ACT
Glu Pro Phe Lys Phe Gly Asp Val Val Phe Asp Asp Pro Ile Thr 291
260 265 270 275
CTC GAC AAC AAC ATT CTT CAC TCT CCT CTC GTC GGA AGA GCG CAG GCA TTT
Leu Asp Asn Asn Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe 339
275 280 285 290
TAT TGG TAC AAC ATG AAG ACT TAC AAC GCT TGG TGG GGC TTC ACA
Tyr Leu Tyr Asn Met Lys Thr Tyr Asn Ala Tyrp Leu Gly Phe Thr 387
295 300 305
TTT GTC CTC AAT TCG ACA GAT TAT AAG GGC ACA ATC ACC TTC ATT GGC
Phe Val Leu Asn Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Asn Gly 435
310 315 320
GCC GAC CCC CTC GTC ATT ARG TAC AGA GAT ATA TCC GTC GTC GGC GGT
Ala Asp Pro Pro Leu Val Lys Tyr Arg Asp Ile Ser Val Val Gly 483
325 330 335
ACG GGT GAT TTC TTG ATG GCG AGA GGA GCA ATT GCC ACC CTC ACT GAT
Thr Gly Asp Phe Leu Met Ala Arg Gly Ile-Ala Ser Thr Asp 531
340 345 350
GCA ATC GAG GGA AAT GAT TAT TTC CGA CTC AGG GTC AAC ATC ACA CTC
Ala Ile Glu Gly Asn Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu 579
355 360 370
TAC GAG TGT TAC TGTGTTAC TRACTAATG GAGAGCTTT GTTTAGAGAA
Tyr Glu Cys Tyr 631

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TATTTGAGT CGAATTTAA AAAATTTAA AAAATTTAA AAAATTTAA AAAATTTAA
AA 871
873
(2) INFORMATION FOR SEQ ID NO:21:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 189 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(iii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
Met Ser Arg Ile Ala Phe His Leu Cys Phe Met Gly Leu Leu Leu Ser 1 5 10 15
Ser Thr Val Leu Arg Asn Val Asp Gly His Ala Thrp Lys Arg Gln Leu 20 25 30
Pro Met Pro Cys Lys Asn Leu Val Leu Tyr Phe His Asp Ile Leu Tyr 35 40 45
Asn Gly Lys Asn Ile His Asn Ala Thr Ala Ala Leu Val Ala Ala Pro 50 55 60
Ala Thrp Gly Asn Leu Thr Thr Phe Ala Glu Pro Phe Lys Phe Gly Asp 65 70 75 80
Val Val Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser 85 90 95
Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Leu Tyr Asn Met Lys Thr 100 105 110
Thr Tyr Asn Ala Thrp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp 115 120 125
Tyr Lys Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Pro Leu Val Lys 130 135 140
Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala 145 150 160
Arg Gly Ile Ala Thr Leu Ser Thr Asp Ala Ile Glu Gly Asn Val Tyr 165 170 175
Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr 180 185
(2) INFORMATION FOR SEQ ID NO:22:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 867 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

TAGTGTGTCG GGCCTTTAAC TTAAAGTCGA CGTTCTATGC AGTGTGAGTC TTGTTTGA
TGATTCGATTTGTTGTTTCTTCCGTT GAGGGTTAAC ATTCACACTCT AGCAGTGTAA
CTGTTATTTG TTATGTTTGGAGAGCTTT GTAACTTGAG ATATAGTAT TTGAGCTTT

691

751

811

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp2 cd

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 80..655

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCAATATGCTCTCTTTGCAAAAATG GCA ATG AAG GCT GCA AAA TTT CTG CAT TTC
TCTCTCTCTTTGCTCTGAAATG GCA ATG AAG GCT GCA AAA TTT CTG CAT TTC
Met Ala Met Lys Ala Ala Lys Phe Leu His Phe 190 200

TAA TTT ATC TGG CTT CTA GTC TGC ACT GTG TGG CTC AAA TCT GCA GAC
Leu Phe Ile Trp Leu Val Cys Thr Val Leu Lys Ser Ala Asp 205 215

TGT CAT AGA TGG AAG AAG AAA ATT CCA GAG CCA TGT AAG ATT CTG GTA
Cys His Arg Trp Lys Ile Pro Glu Pro Cys Lys Asn Leu Val 220 230

TTC TAC TTT CAT GAT ATC CTC TAC TAA GGA TCC AAC AAA CAC AAT GCA
Leu Tyr Phe His Asp Ile Leu Tyr Asn Gly Ser Asn Lys His Asn Ala 235 245

ACA TCT GCA ATT GTT GGA GCA CCC ABA GGA GCC AAT CTC ACT ATT TTG
Thr Ser Ala Ile Val Gly Ala Pro Lys Gly Ala Asn Leu Thr Ile Leu 250 260

ACT GGT AAC AAC CAT TTT GGA GAT GTG GTT GTG TTT GAT CCT ATT
Thr Gly Asn Asn His Phe Gly Asp Val Val Val Phe Asp Asp Pro Ile 265 275

ACT CCT GAC AAC AAT CCT CAC TCT ACT CCT GTG GGA AGA GCT CAG GGC
Thr Leu Asp Asn Asn Ser Thr Pro Val Gly Arg Ala Gln Gly 280 295

TTT TAT TTC TAT GAC ATG AAG AAT ACA TTC AAT TCT TGG CTT GGG TTT
Phe Tyr Phe Tyr Asp Met Lys Asn Thr Phe Asn Ser Trp Leu Gly Phe 300 310

ACA TTT GTG TIG AAT TCA ACA AAT TAT AAG GGC ACC ATC ACC TTC ATT
Thr Phe Val Leu Asn Ser Thr Asn Tyr Lys Gly Thr Ile Thr Phe Asn 315 320

GGG GCT GAC CGA ATT CTG ACT AAG TAC AGA GAT ATA TCT GTT GTC GGT
Gly Ala Asp Pro Ile Ile Thr Lys Tyr Arg Asp Ile Ser Val Gly 330 335

GGT ACG GGT GAT TTC TIG ATG GCC AGA GGA ATC GCC ACC ATT TCT ACT
Gly Thr Gly Asp Phe Ile Met Ala Arg Gly Ile Ala Thr Ile Ser Thr 345 355

GAT GCA TAC GAG GGA GAT GTT TAT TTC CGT CTT AGG GTG AAT ATC ACT 640
 Asp Ala Tyr Glu Gly Asp Val Tyr Phe Arg Leu Asn Ile Thr 375
 370
 375

CTC TAT GAG TGT TAC TGT GTC GAA TGT GTC TCT GTT GTC TCT CTA ATT GAG 695
 Leu Tyr Glu Cys Tyr 380

AGGATGAAACA TTCAATTAAAC TTATAGAG CATATATAA TAGTGCGGG AAAATAAGAG 755

CTAAGGGATG AGATATTAC AGCCCTCATAT CTATATCGC ATCAGTTTG TAGTCATT 815

TGTTATATAA AATTGACCA GTTTCATCAT GTTGAAAGAA AAAAAGAA AA 867

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS: ^
 (A) LENGTH: 192 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Met Ala Met Lys Ala Ala Lys Phe Leu His Phe Leu Ile Trp Leu	1	5	10	15
Leu Val Cys Thr Val Leu Lys Ser Ala Asp Cys His Arg Trp Lys	20	25	30	
Lys Lys Ile Pro Glu Pro Cys Lys Asn Leu Val Leu Tyr Phe His Asp	35	40		
Ile Leu Tyr Asn Gly Ser Asn Lys His Asn Ala Thr Ser Ala Ile Val	50	55	60	
Gly Ala Pro Lys Gly Ala Asn Leu Thr Ile Leu Thr Gly Asn Asn His	65	70	75	80
Phe Gly Asp Val Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn	85	90	95	
Leu His Ser Thr Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp	100	105	110	
Met Lys Asn Thr Phe Asn Ser Trp Leu Gly Phe Thr Phe Val Leu Asn	115	120	125	
Ser Thr Asn Tyr Lys Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile	130	135	140	
Leu Thr Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe	145	150	155	160
Leu Met Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ala Tyr Glu Gly	165			175

Asp Val Tyr Phe Arg Val Arg Val Ile Thr Leu Tyr Glu Cys Tyr
180 185 190 195

(2) INFORMATION FOR SEQ ID NO:24:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 914 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thymidine dirigent protein PSD-Tp3 cDNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 94 .. 669

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CTGAGGAAAT ATCTCTAGGG GAGGCCAAAA TGTGATGTT GTT GTC TCT AAA ACA GCT GCT AGA
 AGATTAGATT CAGGGAAATT TGCAGATGTT GTT GTC TCT AAA ACA GCT GCT AGA
 Val Ser Lys Thr Ala Ala Arg
 195

GTT CTC CTC TTA TGC TTT CTC TGG CTT CTC GTC GCA ATC TTC ATA
 Val Leu His Leu Cys Phe Leu Trp Leu Ile Val Ser Ala Ile Phe Ile
 200 205 210 215

AAA TCT GCA GAT TGC CGT AGC CTT CCA AAG CCC TGT
 Lys Ser Ala Asp Cys Arg Ser Trp Lys Ile Pro Lys Pro Cys
 220 225 230

AGA AAT CTC GTG TTA TAT TTT CAT GAT ATA ATC TAC AAC GGC ADA AAT
 Arg Asn Leu Val Leu Tyr Phe His Asp Ile Ile Tyr Asn Cys Lys Asn
 235 240 245

GCA GAG AAT GCA ACA TCT GCA CTT GTC GCG CCT CAA GGA GCT AAT
 Ala Glu Asn Ala Thr Ser Ala Leu Val Ser Ala Pro Gln Cys Ala Asn
 250 255 260

CTC ACC ATG ACT GGT AAT AAC CAT TTT GGG AAT CTT GCA GTG FTT
 Leu Thr Ile Met Thr Gly Asn Asn His Phe Gln Leu Ala Val Phe
 265 270 275

GAT GAT CCT ATT ACT CTT GAC AAC AAT CTT CAC TCT CCT GTT GGA
 Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Val Gly
 280 285 290

AGA GCT CAG GGC TTT TAC TIC TAT GAC AAG AAC ACC TIC AAT GCC
 Arg Ala Gln Gly Phe Tyr Asp Met Lys Asn Thr Phe Ser Ala
 300 305

TGG CTT GGC TTC ACA TTT GTG CTC AAT GAT CRC AAG GGC TCC
 Trp Leu Gly Phe Thr Val Leu Asn Ser Thr Asp His Lys Gly Ser
 315 320 325
 ATT ACT TTC AAT GGA GCA GAT CCC ATC TPA ACA AAG TAC AGA GAC ATA
 Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr Lys Tyr Asp Ile
 330 335 340
 TCT GTG GGT GGA ACA GGG GAT TGC TNG ATG GCA AGA GGA ATT GCT
 Ser Val Val Gly Gly Ile Thr Gly Asp Leu Met Ala Arg Gly Ile Ala
 345 350 355
 ACC ATT TCT ACT GAC TCA TAT GAG GGA GAT GTC TAT TTC AGG CTT AGG
 Thr Ile Ser Thr Asp Ser Val Tyr Gly Asp Val Tyr Phe Arg Leu Arg
 360 365 370 375
 GTC AAT ATC ACA CTC TAT GAG TGT TAC TGAACGATT CCTTGCTCTG
 Val Asn Ile Thr Ile Val Tyr Glu Cys Tyr
 380
 TATTCTAGT TTGGGACCC TTAAAGAT AGTGTGTTAC TTGAGTGTCT CTATATGTA
 TAACACTGTG TGAGGATTAT ATACAGATGAA CTATAGAAC TATGTGAAAT TCTGTGTCGT
 395
 AGCTAAATTA TGATATGATG COACTCATAT CTCTAAATAT GATAACGATT TGTAAATTAC
 CCAGATAAAG TATGTCATGT GCTTGACAA AAAAARRAAA AAAA
 414

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 192 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Val Ser Lys Thr Ala Ala Arg Val Leu His Leu Cys Phe Leu Trp Leu
 1 5 10 15
 Leu Val Ser Ala Ile Phe Ile Lys Ser Ala Asp Cys Arg Ser Trp Lys
 20 25
 Lys Lys Leu Pro Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp
 35 40
 Ile Ile Tyr Asn Gly Lys Asn Ala Glu Asn Ala Thr Ser Ala Leu Val
 50 55 60
 Ser Ala Pro Gln Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His
 65 70 75 80
 Phe Gly Asn Leu Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn
 85 90 95
 Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp
 100 105 110

Met Lys Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn
115 120 125
Ser Thr Asp His Lys Gly Ser Ile Thr Phe Asn Gly Ala Asp Pro Ile
130 135 140
Leu Thr Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe
145 150 155 160
Leu Met Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ser Tyr Gly Gly
165 170 175
Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Gly Cys Tyr
180 185 190
Met Lys Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn
115 120 125
Ser Thr Asp His Lys Gly Ser Ile Thr Phe Asn Gly Ala Asp Pro Ile
130 135 140
Leu Thr Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe
145 150 155 160
Leu Met Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ser Tyr Gly Gly
165 170 175
Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Gly Cys Tyr
180 185 190

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 704 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp4 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..416

(vi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AG AAT GCC CAC AAT GCA ACA TCT GCA CTT GTC GCA GCC CCT GAG GGA
Asn Ala His Asn Ala Thr Ser Ala Leu Ala Pro Glu Gly Ala
195 200 205
GCC AAT CTC ACC ATT ATG ACT GGT AAT AAC CAT TTT GGG AAT ATT GCT
Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly Asn Ile Ala
210 215 220
GTT TTT GAT GAT CCT ATT CTT GAC AAC AAT CTT CAC TCT CCT TCT
Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Ser Pro Ser
225 230 235
GTT GGA AGA GCT CAG GGC TTT TAC TTC TAT GTC ATT GAG GAT ACC TTC
Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asp Thr Phe
240 245 250
AAT GCT TGG CTT GGT TTT ACA TTT GTG CTG AAT TCA ACT GAT CAC AAG
Asn Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp His Lys
260 265 270
GCC ACC ATT ACT TTC AAT GGA GCA GAT CCA ATC CTG ACC AAG TAC AGA
Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg
275 280 285

GAT ATA TCT GTC GTC GGT GCA ACA GGG GAT TAC TTG ATG GCC AGA GGA
Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg Gly
290 295 300
ATT GCC ACC ATT TCT ACT GAT TCA TAT GAG GGA GAT GTC TAT TAC AGG
Ile Ala Thr Ile Ser Thr Asp Ser Tyr Gly Asp Val Tyr Phe Arg
305 310 315
CTT AGG GTC AAT ATC AGA CTC TAT GAG TGT TAC TAAATGAA TTCCCTCTGT
Leu Arg Val Asn Ile Thr Leu Tyr Gly Cys Tyr
320 325 330
ATTAGTAGCT TATAGGCTC ATTCCTGGT TCAATGCTCA GGGCATGAA TAAAGGATT
TGAAATGCT TTGAAATAT GGAGCATGAA TCTATTTG AAGGCCCTC ARGAGAGTC
ATTTTACAGA GTTGTAGTTT GCCCTCTAGA ATATTATGT TTCATATGC TCTATGAAAG
TCATATGAG TATGGACTAC CATTGGAAAT ATTAAAGCA AGCATATTT ATTAAAGAA
AAAAAAAGAAAAAAGAAAAAAGAAAAA
704
(2) INFORMATION FOR SEQ ID NO:27:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 138 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:
Asn Ala His Asn Ala Thr Ser Ala Leu Val Ala Ala Pro Glu Gly Ala
1 5 10 15
Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly Asn Ile Ala Val
20 25 30
Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Ser Val
35 40 45
Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asp Thr Phe Asn
50 55 60
Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp His Lys Gly
65 70 75
Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg Asp
85 90 95
Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg Gly Ile
100 105 110
Ala Thr Ile Ser Thr Asp Ser Tyr Gly Asp Val Tyr Phe Arg Leu
115 120 125
436

-89-

-90-

Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
130 135

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 820 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: Thujaplicata dirigent protein PSD-Tp5 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 43..612

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GTCTAATGAG GAGAAATTC GAAATTTT TTAACCTAG CA ATG AAA GCC ATT
Met Lys Ala Ile
140

AGA GTT CTC CAT TTA TGC TTT CTA TGT CTT CTA GTG TCT GCA ATC TGC
Arg Val Ile His Leu Cys Leu Cys Leu Leu Val Ser Ala Ile Leu
145 150

CTA AAA TCT GCA GAT TGC CAT AGC TGG AAA AAG TAG CTT CCA AAG CCC
Leu Lys Ser Ala Asp Cys His Ser Trp Lys Lys Leu Pro Lys Pro
160 165

TGC AAG AAT CTT GTG TTA TAT TTC CTT GAT ATA ATC TAC AAT GGC AAA
Cys Lys Asn Leu Val Leu Tyr Phe His Asp Ile Ile Tyr Asn Glu Lys
175 180

AAT GCA GAG AAT GCA ACA TCT GCA CTT GTT GCA GGC CCT GAG GGA GCC
Asn Ala Glu Asn Ala Thr Ser Ala Leu Val Ala Ala Pro Glu Glu Ala
195 200

AAT CTC ACC ATT ATG ACT GGT AAT AAC CAT TTT GGG AAT CTT GCT GTC
Asn Leu Thr Ile Met Thr Glu Asn Asn His Phe Glu Asn Leu Ala Val
210 215

TTT GAT CCT ATT ACT CTT GGC AAC AAT CTC CAC TCT CCT CCT GTC
Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn His Ser Pro Val
225 230

GGA AGA GGT CAG GGA TTT TAC TTC TAT GAC ATG AAG AAC ACC TTC AGT
Gly Arg Ala Gln Gly Phe Tyr Asp Met Lys Asn Thr Phe Ser
240 245

GCT TGG CTT GGC TTC ACA TTT GTG CTT AAT TCA ACT GAT CAC AAG GGC
Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp His Lys Gly
255 260

ACC ATT ACT TTC ATT GCA GCA GAC CCA ATC CTC ACC AAG TAC AGA GAC
Thr Ile Thr Phe Asn Glu Ala Asp Pro Ile Leu Thr Lys Tyr Arg Asp
275 280

ATA TCT GTT GTG GGT GGA ACA GGG GAT TGC TTG ATG GCC AGA GGA ATT
Ile Ser Val Glu Gly Thr Gly Asp Phe Leu Met Ala Arg Glu Ile
290 295

GCC ACC ATT TCT ACT GAT TCA TAT GAG GGA GAA GTT TAT TGC AGG CTT
Ala Thr Ile Ser Thr Asp Ser Thr Asp Ser Ile Tyr Glu Val Tyr Phe Arg Leu
305 310

AGG GTC AAT ATC ACA CTC TAT GAG TGT TAC TGAGCAATG CCTGTCCTCT
Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
320 325

TCCCTCTGTAAG TTCTTGTTT GGGTGCCTT GAGGATAGT TCTGGCTTC AATGTCCTCG
TATGTRGTA CATGGTCATG GGAGTCATT TTGAGGATA TGAGGATAA GTCTCTATAT
ATATATATAT TGAGGAAAT GAGATCTGTT TTGAGTAGCT CTTTCATTC AAAAAAAA
812

AAAAAAAAA
820

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 190 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Lys Ala Ile Arg Val Leu His Leu Cys Leu Leu Val
1 5
Ser Ala Ile Leu Leu Lys Ser Ala Asp Cys His Ser Trp Lys Lys
20 25
Leu Pro Lys Pro Cys Lys Asn Leu Val Leu Tyr His Asp Ile Ile
35 40

Tyr Asn Glu Lys Asn Ala Glu Asn Ala Thr Ser Ala Leu Val Ala Ala
50 55
Pro Glu Glu Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly
65 70
Asn Leu Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His
85 90
Ser Pro Pro Val Gly Arg Ala Glu Gly Phe Tyr His Asp Met Lys
100 105
Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr
115 120

534
582
632
692
752
812
820

-93-

-94-

Leu Pro Lys Pro Cys Lys Asn Leu Val Tyr Phe His Asp Ile Ile 35 40 45
 Tyr Asn Gly Lys Asn Ala Glu Asn Ala Thr Ser Ala Leu Val Ala Ala 50 55 60
 Pro Glu Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly 65 70 75 80
 Asn Leu Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His 85 90 95
 Ser Pro Val Gly Arg Ala Glu Gly Phe Tyr Asp Met Lys 100 105 110
 Asn Thr Phe Ser Ala Tyr Leu Gly Phe Thr Phe Val Leu Asn Ser Thr 115 120 125
 Asp His Lys Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr 130 135 140
 Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met 145 150 155
 Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ser Tyr Glu Gly Asp Val 160 165 170 175
 Tyr Phe Arg Val Asn Ile Thr Leu Tyr Lys Cys Tyr 180 185 190
 (2) INFORMATION FOR SEQ ID NO:32:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 913 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: Thujia plicata dirigent protein P6D-Tp7 . cDNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 71...652
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:
 GCAAGCTCAA ATACCGCACT TCTTCTCTCA CTTCAAGCTC CTTCCTCTT CAACATTT
 TGATGATTT TGCAAG ATG GCA ATC TGG ATG GGA AGA GTT CTG AAC TGC
 Met Ala Ile Tyr Asn Gly Arg Val Leu Asn Leu 109 195 200
 TGC ATT CTG TGG CTT CTC GTC TCC ATA GTT TTG CTG AAT GTT ATA GAT
 Cys Ile Leu Tyr Leu Leu Val Ser Ile Val Leu Leu Asn Gly Ile Asp 205 210

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 192
 (B) TYPE: amino acids
 (D) TOPOLOGY: linear

TGC CAT AGT AGA AAA MAG AAG CTT CCA AAG CCA TGT AGG AAT CTT GTC 205
 Cys His Ser Arg Lys Lys Leu Pro Lys Pro Cys Arg Asn Leu Val 230
 TTG TAT TTT CAT GAT ATT ATC TAC AAT GGT AAA AAT GCA GGC AAT GCA 253
 Leu Tyr Phe His Asp Ile Ile Tyr Asn Gly Lys Asn Ala Gly Asn Ala 245
 ACA TCT ACG CTT GTT GCA GCC CCT CAA GGA GCT AAT CTC ACC ATT ATG 301
 Thr Ser Thr Leu Val Ala Ala Pro Gly Ala Asn Leu Thr Ile Met 250
 255 260 265
 ACT GGC AAT TAC CTT GAT GAT GAT CTT GTC TCT GAT GAT CCT ATT 349
 Thr Gly Asn Tyr His Phe Gly Asp Leu Ser Val Phe Asp Asp Pro Ile 270
 275 280
 ACT GTT GAC AAC AAT CTT CAT TCT CCT CCT GTG GGA AGA GCT CAG GGC 397
 Thr Val Asp Asn Asn His Ser Pro Val Gly Arg Ala Glu Gly 285 290
 TTT TAC TCT TAT GAC ATG AAG AAT ACA TTC AGT GCT TGG CTT GGG TTC 445
 Phe Tyr Phe Asp Met Lys Ser Thr Phe Ser Ala Trp Leu Gly Phe 300 305 310
 ACA TTT GTG CTG AAC TCA GAT TAT AAA GGC ACT ACT TTC GGT 493
 Thr Phe Val Leu Asn Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Gly 315 320
 GGA GCA GAC CCA ATT TTG GCT AAG TAC AGA GAT ATA TCT GTT GTG GGT 541
 Gly Ala Asp Pro Ile Leu Ala Lys Tyr Arg Asp Ile Ser Val Val Gly 330 335 340 345
 GGT ACT GGA GAT TTC TTG ATG GCA AGA GGA ATT GCT ACA ATC GAT ACT 589
 Gly Thr Gly Asp Phe Leu Met Ala Arg Gly Ile Ala Thr Ile Asp Thr 350 355 360
 GAT GCA TAT GAG GGA GAT GTT TAT TTC AGG CTA AGG GTC AAT ATC ACA 637
 Asp Ala Tyr Glu Gly Asp Val Val Phe Arg Leu Arg Val Asn Ile Thr 365 370 375
 CTC TAT GAG TGT TAC TGTATCCATGG GTATTCTATG TAGATACTCT CAATCTGATA 692
 Leu Tyr Glu Cys Tyr 380
 TGGCTATATTT ATTTCGAGAG CATTAGGTAGT TAACTTTAT TAACTAAGTG TGACCATGTA 752
 GATCATGAA AACCTGGGTS CTCATGCGAA GTTTCTAAATA AGTCGCTCG 812
 ACTATACAT TTATGATG TGCGGAAATTG TGTGCGTTAT TACTTTATGA ATATAGCTATT 872
 TTAAACAGAG TTTCACAGG TTAAAAAAA AAAAAAAA A 913

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Ala Ile Trp Asn Gly Arg Val Leu Asn Leu Cys Ile Leu Trp Leu	1	5	10	15
Leu Val Ser Ile Val Leu Asn Gly Ile Asp Cys His Ser Arg Lys	20	25	30	
Lys Lys Leu Pro Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp	35	40	45	
Ile Ile Tyr Asn Gly Lys Asn Ala Gly Asn Ala Thr Ser Thr Leu Val	50	55	60	
Ala Ala Pro Gln Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Tyr His	65	70	75	80
Pho Gly Asp Leu Ser Val Phe Asp Asp Pro Ile Thr Val Asp Asn Asn	85	90	95	
Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp	100	105	110	
Met Lys Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn	115	120	125	
Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Gly Gly Ala Asp Pro Ile	130	135	140	
Leu Ala Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe	145	150	155	160
Leu Met Ala Arg Gly Ile Ala Thr Ile Asp Thr Asp Ala Tyr Glu Gly	165	170	175	
Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr	180	185	190	

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 830 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp8 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 44 .. 619

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CAGAGCTCTT CTTCCTCAA ACATTTGA TATATTGCA ACA ATG GAA ATC TGG	55
Met Ala Ile Trp	
Asn Gly Arg Val Leu Asn Leu Cys Ile Leu Tyr Ile Val Ser Ile	103
Asn Gly Arg Val Leu Asn Leu Cys Ile Leu Tyr Ile Val Ser Ile	210
GTT TTG CTG ATG GGT ATA GAT TGC CAT AGT AGA ABA ARG CTT CCA	151
Val Leu Asn Gly Ile Asp Cys His Ser Arg Lys Lys Leu Pro	220
AAG CCA TGT AGG ATG CTT GTT TGT TAT CAT GAT ATT ATC TAC AAT	199
Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp Ile Tyr Asn	240
GGT AAA ATG GCA GGC ATG GCA ACA TCT ACG CTT GTT GCA GCC CCT CAA	247
Gly Lys Asn Ala Gly Asn Ala Thr Ser Thr Leu Val Ala Pro Gln	255
GGA GCT ATG CTC ACC ATT ATG ACT GCC AAT TAC CAT ATT GGA GAT CTG	295
Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Tyr His Phe Gly Asp Leu	270
GCT GTG TTT GAT GCT ATT ACT GTT GAC AAC AAT CTT CAT TCT CCT	343
Ala Val Phe Asp Asp Pro Ile Thr Val Asn Asn Leu His Ser Pro	285
CCT GTG GCA AGA GCT CAG GGC TTT TAC TTC TAT GAC ATG AAG ATT ACA	391
Pro Val Gly Arg Ala Gln Gly Phe Tyr Asp Met Lys Asn Thr	305
TTC AGT GCT TGG CTT GGG TGC AGA ATT GTG CTG AAC TCA ACA GAT TAT	439
Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp Tyr	320
AAA GGC ACT ATT ACT TTC GGT GGA GCA GAC CCA ATT TTG GCT AAG TAC	487
Lys Gly Thr Ile Thr Phe Gly Ile Asp Pro Ile Leu Ala Lys Tyr	335
AGA GAT ATA TCT GTT GGT ACT GGA GAT TTC TTG ATG GCA AGA	535
Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg	330
(A) LENGTH: 830 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
GTATTCTATG TAGAATAGCT CAACTCGATA TGGCTATATT ATTTCGAGC ATAGGTAGT	629
TAGTTTTAT AACTAAGTAG TGAAACCATGAA GATCATGAA AACCTGGGTG CTCATGCA	749
GTTTCATAT TTCTTAATA AGCTGCTCG ACTATTACAT TTATGGATG TTGAGATTG	809

-97-

TGTCCCTAT TACTTATGAA ATGAGCTT TTAACAGG TTTGACAG TTGAGTT 869
GTCAAAAAA AAAAAGAAA A 890

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met Ala Ile Trp Asn Gly Arg Val Ile Asn Leu Cys Ile Leu Trp Leu: 1 5 10 15
Leu Val Ser Ile Val Leu Asn Gly Ile Asp Cys His Ser Arg Lys 20 25 30
Lys Lys Leu Pro Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp 35 40 45
Ile Ile Tyr Asn Asn Ala Gly Asn Ala Thr Ser Thr Leu Val 50 55 60
Ala Ala Pro Gln Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Tyr His 65 70 75 80
Phe Gly Asp Leu Ala Val Phe Asp Asp Pro Ile Thr Val Asp Asn Asn 85 90 95
Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp 100 105 110
Met Lys Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn 115 120 125
Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Gly Ala Asp Pro Ile 130 135 140
Leu Ala Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe 145 150 160
Leu Met Ala Arg Gly Ile Ala Thr Ile Asp Ala Thr Asp Ala Tyr Gly 165 170 175
Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr 180 185 190

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal sequence from Forsythia intermedia (+)-pinotresinol/(-)-lariciresinol reductase
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Leu Gly Arg 1 5 10 15
Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr 20 25 30

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal tryptic fragment from Forsythia intermedia (+)-pinotresinol/(-)-lariciresinol reductase
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Phe Met Asp Ile Ala Met Xaa Pro Gly Lys Val Thr Leu Asp Glu Lys 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal tryptic fragment from Forsythia intermedia (+)-pinotresinol/(-)-lariciresinol reductase

99-

-100-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Leu Pro Xaa Glu Phe Gly Met Asp Pro Ala Lys Phe Met
 1 5 10

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal tryptic fragment from *Forsythia intermedia* (+)-pinoresinol/(-)-lacticiresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Glu Val Val Gln Xaa Xaa Glu Lys
 1 5

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal tryptic fragment from *Forsythia intermedia* (+)-pinoresinol/(-)-lacticiresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Tyr Xaa Ser Val Glu Glu Tyr Leu Lys Arg
 1 5 10

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 amino acids
 - (B) TYPE: amino acid
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal cyanogen bromide fragment from *Forsythia intermedia* (+)-pinoresinol/(-)-lacticiresinol reductase

- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: internal cyanogen bromide fragment from *Forsythia intermedia* (+)-pinoresinol/(-)-lacticiresinol reductase
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met
 1 5 10

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: internal cyanogen bromide fragment from *Forsythia intermedia* (+)-pinoresinol/(-)-lacticiresinol reductase
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Met Asp Pro Ala Lys Phe Met
 1 5

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: internal cyanogen bromide fragment from *Forsythia intermedia* (+)-pinoresinol/(-)-lacticiresinol reductase
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

-101-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Met Leu Ile Ser Phe Lys Met
 1 5

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: "PCR primer PRL14R"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

ATHATHGGNG GNACNGNTAA
 20

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: "PCR primer PRL14R"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

20

ATTTCGGAC GAGAAACCA GAGAGAG ATG GGA AAA AGC AAA GGT TTG ATC
 Met Gly Lys Ser Lys Val Ile Ile 195

ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA ACT TTA
 Ile Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Lys Ala Ser Leu 205

GCT CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GGT
 Ala Gln Gly His Gln Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val 210
 225

GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT
 Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys Met Gln Gly Ala 235
 240

CAT CTT GTC TGT GCT TCT AAC GAT TTC AAC GTC GTC GAG GCT
 His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser Leu Val Glu Ala 250
 260

GTC AAG CTC GTC GAA GTC GAA ATC AGC GCC ATT TCT GGT CAT ATT
 Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser Gly Val His Ile 265
 270

CGA AGC CAT CAA ATT CTT CTT CAA CTC ARG CTT GTC GAA CCT ATT AAA
 Arg Ser His Gln Ile Ile Leu Gln Leu Lys Leu Val Glu Ala Ile Lys 285
 290

GAG GCT GGA ATT GTC TGT AGA TTT TTA CCA TCT GAG TTT GCA ATG GAT
 Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp 300
 305

CCT GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA AAG GTC ACA CCT
 Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu 310
 315

GAT GAG AAG ATG ATG GTC GAA AGG AAA GCA ATT GAA AAG GCT GGG ATT CCT
 Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro 320
 330

TCYTCNARGN TNCYTTNCC 335

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1060 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: *Forsythia intermedia* cDNA PLR-F11

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 28...963

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

ATTCGGAC GAGAAACCA GAGAGAG ATG GGA AAA AGC AAA GGT TTG ATC
 Met Gly Lys Ser Lys Val Ile Ile 195

ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA ACT TTA
 Ile Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Lys Ala Ser Leu 205

GCT CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GGT
 Ala Gln Gly His Gln Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val 210
 225

GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT
 Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys Met Gln Gly Ala 235
 240

CAT CTT GTC TGT GCT TCT AAC GAT TTC AAC GTC GTC GAG GCT
 His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser Leu Val Glu Ala 250
 260

GTC AAG CTC GTC GAA GTC GAA ATC AGC GCC ATT TCT GGT CAT ATT
 Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser Gly Val His Ile 265
 270

CGA AGC CAT CAA ATT CTT CTT CAA CTC ARG CTT GTC GAA CCT ATT AAA
 Arg Ser His Gln Ile Ile Leu Gln Leu Lys Leu Val Glu Ala Ile Lys 285
 290

GAG GCT GGA ATT GTC TGT AGA TTT TTA CCA TCT GAG TTT GCA ATG GAT
 Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp 300
 305

CCT GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA AAG GTC ACA CCT
 Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu 310
 315

GAT GAG AAG ATG ATG GTC GAA AGG AAA GCA ATT GAA AAG GCT GGG ATT CCT
 Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro 320
 330

TCYTCNARGN TNCYTTNCC 335

20

-102-

-103-

TTC ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT TCC TTG GGA GGT
Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly 531
345 350 355 360

CTC TGT CAA TTT GGC AAA ATT CTT CCT CCT TCT AGA GAT TTT GTC ATT ATA
Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile Ile 579
365 370 375

CAT GGA GAT GGT AAC AAA GCA ATA TAT AAC ATT GAT GAA GAT GAT ATA
His Gly Asp Gly Asn Lys Lys Ala Ile Tyr Asn Asn Glu Asp Asp Ile 627
380 385 390 395

GCA ACT TAT GCC ATC ACA ATT ATT GAT CCA AGA ACC CTC AAC AAG
Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys 675
395 400 405

ACA ATC TAC ATT AGT CCT CCA ADA AAC ATC CTT TCA CAA AGA GAA GTT
Thr Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser Gln Arg Gln Val 723
410 415 420

GTT CAG ACA TGG GAG ATT GGG AAA GAA CTG CAG AAA ATT ACA
Val Gln Thr Trp Glu Lys Ile Gly Lys Glu Leu Gln Lys Ile Thr 771
425 430 435 440

CTC TCG AAC GAA GAT TTT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT
Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Gln Ala 819
445 450 455

CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC
Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Gln Gly Cys 867
460 465 470

CTT ACG AGT TTT GAG ATA GGA GAT GAA GCA TCT AAA ATT CTT TAT
Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Ala Ser Asp Leu Tyr 915
475 480 485

CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC AAC CTT TAC GTG
Pro Glu Val Lys Tyr Thr Ser Val Gln Glu Tyr Leu Lys Arg Tyr Val 963
490 495 500

TAGTTGAG CTTTCCATTAA TTTATGTTATATATTTAAAT TCACTATGTA GTTTTAATT
TCTGTTAATA ATATGTTG AATTGTTG CCAAAA 1023

1060

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Met Gly Lys Ser Lys Val Leu Ile Ile Gly Thr Gly Tyr Leu Gly
1 5 10 15

-104-

Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
20 25
Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
35 40 45

Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
50 55 60

Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
65 70 75 80

Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
85 90 95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
100 105 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
115 120 125

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
130 135 140

Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
145 150 155 160

Phe Ala Gly Tyr Phe Leu Gly Leu Cys Gln Phe Gly Lys Ile Leu
165 170 175

Pro Ser Arg Asp Phe Val Ile Ile His Gly Asp Gly Asn Lys Lys Ala
180 185 190

Ile Tyr Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
195 200 205

Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
210 215 220

Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile
225 230 235 240

Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
245 250 255

Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
260 265 270

His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp,
275 280 285

Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
290 295 300

Glu Glu Tyr Leu Lys Arg Tyr Val
305 310

(2) INFORMATION FOR SEQ ID NO:49:

-105-

-106-

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1112 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: *Escherichia coli* cDNA BLR-E12

(iv) HYPOTHETICAL: NO

(v) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 44..979

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```

AATTCGGCAC GAGCTCGGCC CGCAGAGA AAAACAGAGA GAG ATG GGA AAA ACC
          55
          Met Gly Val Ser
AAA GTT TGG ATC ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TGG GTT
          103
          Lys Val Leu Ile Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val
          320
AAG GCA AGT TTA GCT CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT
          151
          Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro
          335
GAA ATT GCT GTT GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA
          199
          Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys
          350
ATG CAA GGA GCT CAT CTT GAA TCT GGT TCT TAC AGG GAT TTC AAC AGT
          247
          Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser
          365
CTG GTC GAG GCT AAC GTC GTC GAC GTC GAA GTC AGC GCC ATT TCT
          295
          Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser
          385
GCT GTT CTT ATT CGA AGC CAT CAA ATT CTT CTT CAA CTC AAC CTT GTT
          343
          Gly Val His Ile Arg Ser His Glu Ile Leu Glu Leu Lys Leu Val
          405
GAA GCT ATT AAA GAG GCT GGA AAT GTC AAC AGC AGN TTT TTA CCA TCT GAG
          391
          Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu
          415
          425
TTT GGA ATG GAT CCT GCA ARA TTT ATG GAT AGC GCC ATG GAA CCC GGA
          439
          Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly
          430
          440
ARG GTC AGA CTT GAT GAG ATG GTG GTC AGG AAG GCA ATT GAA RAG
          487
          Lys Val Thr Leu Asp Glu Lys Met Val Arg Lys Ala Ile Glu Lys
          445
          455
          460

```

```

GCT GGG ATT CCT TTC ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT GAT
          535
          Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr
          465
          470
TTC TGG CGA GGT CTC TGT CAA TTT GGC AAA ATT CTT CCT TCT AGA GAT
          583
          Phe Leu Gly Gly Leu Cys Glu Phe Gly Lys Ile Leu Pro Ser Arg Asp
          480
          490
TTT GTC ATT ATA CAT GGA GRT GGT AAC ARA ARA GCA ATA TAT AAC ATT
          631
          Phe Val Ile His Gly Asp Gly Asn Lys Ala Ile Tyr Asn Asn
          495
          500
          505
GAA GAT GAT ATA GCA ACT TAT GGC ATC ARA AGC ATT AAT GAT CCA AGA
          679
          Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg
          510
          515
          520
ACC CTC AAC AAG ACA ATC TAC ATT AGT CCT CCA AAA AAC ATC CTT TCA
          727
          Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Lys Asn Ile Leu Ser
          525
          530
          535
          540
CAA AGA GAA GTT GTC CAG ACA TGG GAG ARG CTT ATT GGG AAA GAA CTG
          775
          Gln Arg Glu Val Val Gln Thr Trp Glu Ile Gly Lys Glu Leu
          545
          550
          555
CAG AAA ATT ACA CTC TCG AAG GAA GAT TTT TTA GCC TCC GTC AAA GAG
          823
          Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu
          560
          565
          570
CTC GAG TAT GCT CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC
          871
          Leu Glu Tyr Ala Glu Gln Val Gly Leu Ser His Tyr His Asp Val Asn
          575
          580
          585
TAT CAG GGA TGC CTT AGC AGT ATT GAG ATA GGA GAT GAA GAA GAG GCA
          919
          Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Ala
          590
          600
TCT AAA CCT TAT CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC
          967
          Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu
          605
          610
          620
AAG CGT TAC GTG TAGTGAAG CTTTCATTA TTTATTGTAAT AATATTAA
          1019
          Lys Arg Tyr Val

```

TCAGTATGTA GTTGTAAATT TCGTTAAATAA ATATGTGTTG AATTTGCTT CAAACGAGTG
GTGCGATGAA ATGGGATTTTT GAAAGTCAAAAA AAA

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

-107-

-108-

Met Gly Lys Ser Lys Val Leu Ile Gly Ile Gly Thr Gly Tyr Leu Gly
 1 5 10 15
 Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
 20 25 30
 Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
 35 40 45
 Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
 50 55 60
 Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Ile
 65 70 75 80
 Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
 85 90 95
 Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
 100 105 110
 Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
 115 120
 Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
 130 135 140
 Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
 145 150 155 160
 Phe Ala Gly Tyr Phe Leu Gly Leu Cys Gln Phe Gly Lys Ile Leu
 165 170 175
 Pro Ser Arg Asp Phe Val Ile His Gly Asp Gly Asn Lys Lys Ala
 180 185 190
 Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
 195 200 205
 Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ser Pro Pro Lys
 210 215 220
 Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile
 225 230 235 240
 Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
 245 250 255
 Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
 260 265 270
 His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
 275 280 285
 Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
 290 295 300
 Glu Glu Tyr Leu Lys Arg Tyr Val
 305 310

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1124 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: *Forsythia intermedia* cDNA PLR-F13
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 29..966
- (vi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

```

AATTCGGCAC GAGGAAACAGAGAGATG GGA AAA AGC AAA GTT TTG ATC
      315   Lys Ser Lys Val Ile Ile
      320
ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA AGT TTA
      325   Ile Gly Gln Gln
      330
GCT CAA GCT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GGT
      335   Ala Gln Gln
      340
GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT
      345   Asp Lys Val Glu Met Ile Ser Phe Lys Met Gln Gln Gln Gln
      350
CAT CTT GTC TCT GGT TCT TGC AAG GAT TTC AAC GTC GTC GAG GCT
      355   His Leu Val Ser Gly Ser Phe Asp Ser Leu Val Glu Ala
      360
GTC AAG CTC GTC GAC GTC GTC
      365   Val Lys Leu Val Asp Val Val Asp Val Val Asp Val Val Asp Val
      370
GCA AGC CAA ATT CTT CTT CAA CTC AAG CTT GTC GAA GCT ATT AAA
      375   Arg Ser His Gln Ile Leu Gln Leu Gln Ile Ser Gly Val Ile Lys
      380
GAG GCT GGA AAT GTC AAG AGA TTT TCA TCT GAG TTT GGA ATG GAT
      385   Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp
      390
CCT GCA AAA TTT ATG GAT AGC GCC ATT GCA CCC CGA AAC GTC AGA CCT
      395   Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Glu Lys Val Thr Leu
      400
GAT GAG AAG ATG GTC GTC AGG AAA GCA ATT GAA MAG GCT ATT CCT
      405   Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Ile Gly Ile Pro
      410
      415
      420
      425
      430
      435
      440
      445
      450
      460

```

-109-

-110-

TTC ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT TGC TAC TGC GGA GGT
Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly 532
465 470 475 480

CTC TGT CAA TTT GGC AAA ATT CTT CCT TCT AGA GAT TTT GTC ATT ATA
Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile 580
485 490 495

CAT GGA GAT GGT AAC ARA ARA GCA ATA ATA ARC AAT GAA GAT GAT ATA
His Gly Asp Gly Asn Lys Ile Tyr Asn Asn Gln Asp Asp Ile 628
500 505 510

GCA ACT TAT GCC ATC ARA ARA ATT ATT ATT GAT CCA AGA ACC CTC AAC AAG
Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys 676
515 520 525

ACA ATC TAC ATT AGT CCT CCA AAA AAC ATC CTT TCA CAR AGA GAA GTC
Thr Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser Gln Arg Glu Val 724
530 535 540

GTT CAG ACA TGG GAG BAG CTT ATT GGG ABA GAA CTG CAG BAA ATT ACA
Val Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu Gln Ile Thr 772
545 550 555 560

CTC TCG AAG GAA GAT TTT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT
Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Gln Tyr Ala 820
565 570 575

CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC
Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr 580 585 590 595

CTT ACG AGT TTT GAG ATA GGA GAT GAA GAG GCA TCT AAA CTT TAT
Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Ala Ser Lys Leu Tyr 605
610 615 620

CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC AAC CGT TAC GTG
Pro Glu Val Lys Tyr Thr Ser Val Glu Tyr Leu Lys Arg Tyr Val 964
610 615 620

TAGTGAAAG CTTCCATTAA TTATGTAAT AATATTAAAG TCAGTATGTA GTTTAAATT
TCTGTTAAATA ATATGTTGTTG ATTTCGTTGCTT CAACAGTG GTCGATTGAA ATGGGATTT
GAAGTCATCT TCTCCACATAT ATTAGTCCAA ATAAAGAAA 1024
1084
1124

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Met Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly
1 5 10 15
Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
20 25
Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
35 40 45
Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
50 55 60
Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
65 70 75 80
Ser Ala Ile Ser Gly Val His Ile Arg Ser His Glu Ile Leu Leu Gln
85 90 95
Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Arg Phe
100 105 110
Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
115 120 125
Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
130 135 140
Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
145 150 160
Phe Ala Gly Tyr Phe Leu Gly Gln Phe Gly Lys Ile Leu
165 170 175
Pro Ser Arg Asp Phe Val Ile His Gly Asp Gly Asn Lys Lys Ala
180 185 190
Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
195 200 205
Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
210 215 220
Asn Ile Leu Ser Gln Arg Glu Val Val Glu Lys Leu Ile
225 230 240
Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
245 250 255
Ser Val Lys Glu Leu Glu Tyr Ala Gln Val Gly Leu Ser His Tyr
260 265 270
His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
275 280 285
Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
290 295 300
Glu Glu Tyr Leu Lys Arg Tyr Val
305 310

-111-

.112-

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1097 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: *Forsythia intermedia* cDNA PLR-F14

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 29..964

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

AAATGGCAC GAGGAAAC AGAGAGAG ATG GGA AAA AGC AAA GTT TTG ATC
Met Gly Iys Ser Lys Val Leu Ile 315

ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AGG GCN AGT TTA
Ile Gly Thr Gly Tyr Leu Gly Arg Leu Val Ile 335

GCT CAA GGT CAT GAA ACA TAC ATT CTC CAT AGG CCT GAA ATT GGT GTT
Ala Gln Gly His Glu Thr Tyr Ile His Arg Pro Glu Ile Gly Val 340

GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT
Asp Ile Asp Iys Val Glu Met Leu Ile Ser Phe Lys Met Glu Ala 355

CAT CTT GTC TCT GGT TTC ATG GTC AAC AGT CGT CTC GTC GCG GCT
His Leu Val Ser Gly Ser Phe Lys Asp Ser Asn Ser Leu Val Glu 365

GTC AGG CTC GTC GAC GTC GAA ATC AGC GCC ATT TCT GGT CAT ATT
Val Lys Leu Val Asp Val Ile Ser Ala Ile Ser Gly Val His Ile 370

GAG AGC CAT CAA ATT CTT CTC CAA CTC AAG CTT GTC GAA GCT ATT AAA
Arg Ser His Gln Ile Leu Leu Lys Leu Val Glu Ala Ile Lys 385

GAG GCT GGA ATT GTC ATG AGA TTT TTA CCA TCT GAG TTT GGA ATG GAT
Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp 390

CCT GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA AAG GTC ACA CTT
Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu 405

GAT GAG ATG ATG GTG GTC GAA AGG PAA GCA ATT GAA ATG GCT GGG ATT CCT
Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Ala Gly Ile Pro 410

415 445
455 484
460

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1097 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: *Forsythia intermedia* cDNA PLR-F14

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 29..964

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TTC AGC ATT GTC TCT GCA ATT TGC ATT GCT GAT TTT GTC ATT GGA GGT
Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly Ile 465 475

CTC TGT GAA TTT GGC AAA ATT CTT CCT TCT AGA GAT TTT GTC ATT ATA
Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile Ile 485 490

CAT GGA GAT GGT AAC AAA GCA ATA TAT AAC ATT GAA GAT GAT ATA
His Gly Asp Gly Asn Lys Lys Ile Tyr Asn Asn Glu Asp Asp Ile 500 505

GCA ACT ATT GTC ATC AAA ATT ATT GAT GTC CCA AGA ACC CTC ARC ARG
Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys 515 525

ACA ATC TAC ATT AGT CCT CCA AAA AAC ATC ATT TCA CAA AGA GAA GTC
Thr Ile Tyr Ile Ser Pro Lys Asn Ile Leu Ser Gln Arg Glu Val 530 535

GTT CAG AGC TGG GAG ARG CTT ATT GGG AAA GAA CTG CAG AAA ATT ACA
Val Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu Lys Ile Thr 540 550

CTC TCG AGG GAA GAT TTT TTA GCC TCC GTG AGA GAG -CTC GAG TAT GCT
Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Glu Tyr Ala 555 560

CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC
Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Glu Gly Cys 570 575

CTC AGC AGT TTT GAG ATA GGA GAT GAA GAA GAG GCA TCT AAA CCT TAT
Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Ala Ser Lys Leu Tyr 590 595

CCA GAG GTT AGC ACT GTG GAA GAG GTC TAC CTC AAG CCT TAC TGT
Pro Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu Lys Arg Tyr Val 600 605

TAGTTGAAG CTTTCCATTAA TTATTTAAAT TTATGTAATTA TCAGTATGTA GTTTAAATT
TCGTTAATAA ATATGTTGTT AATTGTTGTT CAACAGAGTG GTCCATTGAA ATGGATTTT
GAAAAAAA ARA 1024
1084
1097

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Met Gly Lys Ser Lys Val Leu Ile Gly Lys Thr Gly Tyr Leu Gly
 1 5 10 15
 Arg Arg Leu Val Lys Ala Ser Leu Ala Glu His Glu Thr Tyr Ile
 20 25 30
 Leu His Arg Pro Glu Ile Gly Ala His Leu Val Ser Gly Ser Phe Lys
 35 40 45
 Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Val Ser Gly Ser Phe Lys
 50 55 60
 Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
 65 70 75 80
 Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
 85 90 95
 Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
 100 105 110
 Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
 115 120 125
 Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
 130 135 140 145
 Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Val Tyr Val Ser Ala Asn Cys
 145 150 155 160
 Phe Ala Gly Tyr Phe Leu Gly Leu Cys Gln Phe Gly Lys Ile Leu
 165 170 175
 Pro Ser Arg Asp Phe Val Ile His Gly Asp Gly Asn Lys Iys Ala
 180 185 190 195
 Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
 200 205 210
 Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
 215 220
 Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Ile Ile
 225 230 235 240
 Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
 245 250 255
 Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Ieu Ser His Tyr
 260 265 270 275
 His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
 280 285 290 295
 Glu Glu Glu Ala Ser Lys Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
 305 310

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1109 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: *Forsythia intermedia* cDNA PLR-Fis

(iv) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 31..966

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATTGGCAC GAGGAGAAA ACAGAGAGAG ATG GGA AAA AGC ARA GTT TTG ATC
 54
 Met Gly Lys Ser Lys Val Leu Ile Gly Val Val Val Ile
 315 320
 ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA AGT TTA
 102
 Ile Gly Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Lys Ala Ser Leu
 325 330 335
 GCT CAA GGT CAT GAA GCA TAC ATT CTG CAT AGG CCT GAA ATT GGT GTT
 150
 Ala Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val
 340 345 350
 GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT
 198
 Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys Met Gln Gly Ala
 355 360 365
 GTC AAG CTC GTC GAC GCA GTC ATA ATC AGC GCC ATT TCT GTT CAT ATT
 294
 Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser Gly Val His Ile
 385 390 395
 His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser Leu Val Glu Ala
 370 375 380
 CGA AGC CAT CRA ATT CTT CTT CAA CTC AAG CTT GTT GAA GCT ATT AAA
 342
 Arg Ser His Gln Ile Leu Leu Gln Leu Lys Leu Val Glu Ala Ile Lys
 405 410 415
 Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu
 420 425 430
 Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp
 440 445 450
 CCT GCA AAA TTT ATG GAT AGC GCC ATG GAA CCC GGA AAG GTA ACA CCT
 438
 GAG GCT GCA ATT GTC AAG AGA TTT TTA CCA TCT GAG TTT GGA ATG GAT
 390
 Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp
 425 430
 GAT GAG ATG GTG GTA AGG AAA GCA ATT GAA ARG GCT GGG ATT CCT
 486
 Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro
 450 455 460

-115-

.116-

TTC AGA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT TTC TTG GGA GGT
 Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly 534
 465 470 475 480

CTC TGT CAA TTT GGC AAA ATT CCT CCT TCT AGA GAT TTT GTC ATT ATA
 Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile 582
 485 490 495

CAT GGA GAT GGT AAC AAA GCA ATA TAT AAC AAT GAA GAT GAT ATA
 His Gly Asp Gly Asn Lys Lys Ile Asn Cys Pro Asp Asp Ile 630
 500 505 510

GCA ACT TAT GCC ATC AAA ACA ATT ATT GAT CCA AGA ACC CTC AAC Arg
 Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys 678
 515 520 525

ACA ATC TAC ATT AGT CCT CCA ARA AAC ATC CTT TCA CAA AGA GAA GTT
 Thr Ile Tyr Ile Ser Pro Pro Asn Ile Leu Ser Gln Arg Glu Val 726
 530 535 540

GTT CAG ACA TGG GAG AAG CTT ATT GGG AAA GAA CTTG CAG AAA ATT ACA
 Val Gln Thr Trp Glu Lys Ile Gly Lys Glu Leu Gln Lys Ile Thr 774
 545 550 555 560

CTC TCG RAG GAA GAT TTT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT
 Leu Ser Lys Gln Asp Phe Leu Ala Ser Val Lys Glu Leu Gln Tyr Ala 822
 565 570 575

CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC
 Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Glu Gln Cys 870
 580 585 590

CTT ACG AGT TTT GAG ATG GAA GAT GAA GAA GRG GCA TCT AAA CTT TAT
 Leu Thr Ser Phe Glu Ile Gly Asp Glu Gln Gln Ala Ser Lys Leu Tyr 918
 595 600 605

CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TGC CTC AAC CGT TAC GTG
 Pro Glu Val Lys Tyr Thr Ser Val Glu Gln Tyr Ile Lys Arg Tyr Val 966
 610 615 620

TAGTGAAGG CTTCCATTA TATTTAAT ATATTATA TCACTATGTA GTTTAAATT
 TCGTTAAATA ATATGTTG TATTTGCTT CAACGAGTG GTCTGATGAA AGGAGTTT 1026

GAAGTCATCT TCTCCAAAAA AAA 1109

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Met Gly Lys Ser Lys Val Leu Ile Ile Gly Thr Gly Tyr Leu Gly
 1 5 10 15
 Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
 20 25 30

Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
 35 40 45

Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
 50 55 60

Asp Phe Asn Ser Leu Val Gln Ala Val Lys Leu Val Asp Val Val Val Ile
 65 70 75 80

Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gin Ile Leu Leu Gin
 85 90 95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
 100 105 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
 115 120 125

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
 130 135 140

Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Val Ser Ala Asn Cys
 145 150 155 160

Pho Ala Gly Tyr Phe Leu Gly Leu Cys Gin Phe Gly Lys Ile Leu
 165 170 175

Pro Ser Arg Asp Phe Val Ile His Gly Asp Gly Asn Lys Lys Ala
 180 185 190

Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
 195 200 205

Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
 210 215 220

Asn Ile Leu Ser Gln Arg Glu Val Val Glu Thr Ile Lys Leu Ile
 225 230 235 240

Gly Lys Glu Leu Gln Iys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
 245 250 255

Ser Val Lys Glu Leu Glu Tyr Ala Gln Val Gly Leu Ser His Tyr
 260 265 270

His Asp Val Asn Tyr Gln Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
 275 280 285

Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
 290 295 300

Glu Glu Tyr Leu Lys Arg Tyr Val
 305 310

-117-

-118-

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1107 bases pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: *Forsythia intermedia* cDNA PLR-F16

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 27..962

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

AATTGGCAC GAGAAACAG AGAGAG ATG GGA ARA NGC AAA GTT TGG ATC ATT
Met Gly Lys Ser Lys Val Leu Ile Ile
315 320

GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AGG GCA AGT TPA GCT
Gly Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Ile Ser Leu Ala
325 330

CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GTT GAT
Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val Asp
340 345

ATT GAT ARA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT CAT
Ile Asp-Lys Val Glu Met Leu Ile Ser Phe Lys Ala Ser Met Glu Gly-Ala His
355 360

CTT GTC TCT GGT TCT TTC AAC GAT TTC AAC AGT Phe Asn Ser Leu Val Glu Ala Val
370 375

AAG CTC GTC GAC GCA GTC ATC AGC GCC ATT TCT GGT CAT ATT CGA
Lys Leu Val Asp Val Ile Ser Ala Ile Ser Gly Val His Ile Arg
390 395

AGC CAT CTC ATA ATT CTT CTC CAA CTC AAC CTT GTC GAA GCT ATT AAA GAG
Ser His Glu Ile Leu Leu Glu Leu Ile Lys Leu Val Glu Ala Ile Lys Glu
405 410

GCT GGA ATT GTC AAG AGA ATT TTA CCA TCT GAG TTT GGA ARG GAT CCT
Ala Gly Asn Val Lys Arg Phe Ile Pro Ser Glu Phe Gly Met Asp Pro
420 425

GCA AAA ATT ATT ATG GAT AGC GCG ATG GAA CCC GGA AGAG GAA AGC CTT GAT
Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu Asp
435 440

GAG AAG ATG GTG GTC AGG AAA GCA ATT GAA AAG CCT GGG ATT CCT TTC
Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro Phe
450 455

ACA TAT GTC TCT GCA AAT TGC ATT GCT GGT ATT TTC TTT GGA GGT CTC
Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly Glu Leu533
470TGT CTA ATT GTC GGC AAA ATT CCT CCT TCT AGA GAT ATT GTC ATT ATA CAT
Cys Gin Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile His581
485GGA GAT GGT AAC AAA GCA ATT GAA ATT GAT GAT ATA GCA
Gly Asp Gly Asn Lys Lys Ala Ile Tyr Asn Asn Glu Asp Asp Ile Ala629
505ACT ATT GCC ATC AAA ACA ATT ATT GAT CCA AGA ACC CTC AAC AAG ACA
Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys Thr677
515ATC TAC ATT AGT CCT CCA AAA AAC ATC ATT CTT TCA GAA AGA GAA ATT GTT
Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser Gin Arg Glu Val Val725
530CAG AGA TGG GAG AAG CCT ATT GGG ARA GAA CTC CAG AAA ATT ACA CTC
Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu Glu Lys Ile Thr Leu773
550TCG AAG GAA GAT ATT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT CAG
Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Tyr Ala Glu821
565CAG GTG GCA TTA AGC CTT ATT CAT GAT GTC AAC TAT CAG GCA TGC CTC
Gin Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Glu Cys Leu869
580ACG AGT ATT GAG ATA GCA ATT GAA GAG GCA TCT AAA CTT TAT CCA
Thr Ser Phe Glu Ile Gly Asp Glu Glu Ala Ser Lys Leu Tyr Pro917
595GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC AAG CGT TAC GTC
Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu Lys Arg Tyr Val962
610TAGTGAAAGG CTTCCTTATTATTATTAAATTTTTAAATTTTTAAATT
TCAGTAGTAAATAATATGTGTTGAAATTTGCTTCAAAAGGAGTGGATGGATT
GAGTCATC TCTCCTGAA AAAA1022
615

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Met Gly Lys Ser Lys Val Leu Ile Ile Gly Glu Thr Gly Tyr Leu Gly
 1 5 10 15
 Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
 20 25 30
 Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
 35 40 45
 Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
 50 55 60
 Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
 65 70 75 80
 Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
 85 90 95
 Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
 100 105
 Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
 115 120 125
 Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
 130 135 140
 Ala Ile Glu Lys Ala GLY Ile Pro Phe Thr Val Ser Ala Asn Cys
 145 150 155 160
 Phe Ala Gly Tyr Phe Leu Gly Leu Cys Gln Phe Gly Lys Ile Leu
 165 170 175
 Pro Ser Arg Asp Phe Val Ile Ile His Gly Asp Gly Asn Lys Ala
 180 185 190
 Ile Tyr Asp Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
 195 200
 Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
 210 215 220
 Asn Ile Leu Ser Gln Arg Glu Val Val Glu Thr Trp Glu Lys Leu Ile
 225 230 235 240
 Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
 245 250 255
 Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
 260 265 270
 His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
 275 280 285
 Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
 290 295
 Glu Glu Tyr Leu Lys Arg Tyr Val
 310 315

(2) INFORMATION FOR SEQ ID NO:59:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: "cDNA synthesis linker primer"
 (iii) HYPOTHETICAL: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:
 GTCTCGAGTT TTTTTTTT TTTTTT
 (2) INFORMATION FOR SEQ ID NO:60:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: "cDNA synthesis primer"
 (iii) HYPOTHETICAL: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:
 GCACATAGAA GTATGATAA G
 (2) INFORMATION FOR SEQ ID NO:61:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1190 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: Thymidine kinase cDNA PR-1p1
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 13..951
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:
 GCACATAGAA GT ATG GAT AAG AAG AGC AGA GTT CTG ATA GTC GGG GGC
 Met Asp Lys Ser Arg Val Leu Ile Val Gly Gly
 315 320

-121-

-122-

ACT GGT TAT ATA GGC AAA AGA ATT GTC ATT GCC AGT ATA TCT CTT GGC
Thr Gly Tyr Ile Gly Lys Arg Ile Val Asn Ala Ser Ile Ser Leu Gly 96
325 330 335 340 344 350 355 360 365 370

GAT CCC ACT TAT GTC ATT GTC AGA CCA GAA GTG GTC TCT AAC ATT GAC
His Pro Thr Tyr Val Leu Phe Arg Pro Glu Val Val Ser Asn Ile Asp 345 350 355 360 365 370 375 380 385 390 395

AAA GTG CAG ATG CTG TTA TAC TTC AAA CAG CTT GGT GCC AAA CTT ATT
Lys Val Gln Met Leu Phe Lys Glu Leu Gly Ala Lys Leu Ile 360 365 370

GAG GCT TCA TTC GAT GAC CAC CAA AGG CTT GTG GAT GCT CTG AAA CAA
Glu Ala Ser Leu Asp Asp His Glu Arg Leu Val Asp Ala Leu Lys Gln 380 385 390 395 400

GTG GAT GTT GTC ATA AGT GCT TGT GCA GGA GGT GTT CTA AGC CAC CAT
Val Asp Val Val Ile Ser Ala Leu Ala Gly Val Leu Ser His His 390 395 400 405 410 415 420

ATA CTT GAA CAG CTC BAA CTA GTG GRA GCC ATC AAA GCA GCT GGA AAT
Ile Leu Glu Gln Leu Iys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn 405 410 415 420

ATT AAG AGA TTT CTT CCA TCT GAG TTT GGC ATG GAT ATT ATG
Ile Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp Pro Asp Ile Met 425 430 435

GAG CAT GCA TGG CAA CCT GGT AGC ATT ACA TTC ATC GAT AAG AGA AAG
Glu His Ala Leu Gln Pro Gly Ser Ile Thr Phe Ile Asp Lys Arg Lys 440 445 450

GTT CGG CGT GCC ATT GAA GCA TCC ATT CCT TAC TAC ATA TAT GTG TCT
Val Arg Arg Ala Ile Glu Ala Ala Ser Ile Pro Tyr Ile Val Ser 455 460 465 470 475 480

TCA AAT ATG TTT GCT GGT TAC TTT GCT GGA AGT TTA GCT CAA CTT GAT
Ser Asn Met Pro Phe Ala Gly Tyr Phe Ala Gly Ser Leu Ala Gln Leu Asp 470 475 480 485 490 495 500

GGT CAT ATG ATG CCT CCT CGA GAC AGG GTC CTC ATC TAT GGA GAT GGA
Gly His Met Met Pro Pro Arg Asp Iys Val Leu Ile Tyr Gly Asp GLY 485 490 495 500

AAT GTT AAA GGT ATT TGG GTG GAT GAA GAT GAT GAT GTC GGA ACA TAC ACA
Asn Val Lys Gly Ile Ile Val Asp Glu Asp Val Gly Thr Tyr Thr 505 510 515

ATC AAA TCA ATT GAT GAT CCA CAA ACC CTT AAC AAG ACT ATG TAT ATT
Ile Lys Ser Ile Asp Asp Pro Glu Thr Leu Asn Lys Thr Met Tyr Ile 520 525 530 535 540 545

AGG CCA CCT ATG AAT ATC CTT TCA CAG AAG GAA GGT ATA CAA ATA TGG
Arg Pro Pro Met Asn Ile Ile Ser Glu Lys Glu Val Ile Gln Ile Trp 545 550 555 560

GAG AGA TTA TCA GAA CAA CAC ATC CTG GAT AAA ATA TAC ATT TCT TCT CAA
Glu Arg Leu Ser Glu Gln Asn Leu Asp Lys Ile Tyr Ile Ser Ser Gln 555 560

GAC TTT CTT GCA GAT ATG AAA GAT AAA TCA TAT GAA GAG ARG ATT GTA
Asp Phe Leu Ala Asp Met Lys Ser Tyr Glu Glu Lys Ile Val 565 570 575 580 585 590 595

CGA TGT CAT CTC TAC CAA ATT TTC TTT AGA GGA GAT CTT TAC AAC TTT
Arg Cys His Leu Tyr Glu Ile Phe Phe Arg Gly Asp Leu Tyr Asn Phe 595 600 605 610 615 620 625

GAA ATT GCC CCC ATT CCT ATT GAA GCT ACC AAA CTT TAT CCA GAA GTG
Glu Ile Gly Pro Asn Ala Ile Glu Ala Thr Lys Leu Tyr Pro Glu Val 605 610 615 620 625 630 635

AAA TAC GTC ACC ATG GAT TCA TAT TTA GAG CGC TAT GTT TGATATCTT
Lys Tyr Val Thr Met Asp Ser Tyr Leu Glu Arg Tyr Val 615 620 625 630 635 640 645

TCTAGTTTG TATATGTTT TTCTCATGAA TAATGTGAGA GGTCATATTG CAAATAATT
AGCTATGG CTGAAATTAA AACTAGACT ACACCTTTT CCAAAATTACT TACATCTATT 650 655 660 665 670 675 680

TTTACTCAT ATTTGACTCTA ATTTGACTCTA ATTTGACTCTA ATTTGAACTC ATTAATGATAT
TATAATTAT TATAGATCTT ATTAAATAATA AAAAAAAA AAAAAAAA 685 690 695 700 705 710 715

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 313 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Met Asp Lys Ser Arg Val Leu Ile Val Gly Glu Thr Gly Tyr Ile 1 5 10 15

Gly Lys Arg Ile Val Asn Ala Ser Ile Ser Leu Gly His Pro Thr Tyr 20 25 30

Val Leu Phe Arg Pro Glu Val Val Ser Asn Ile Asp Lys Val Glu Met 35 40 45

Leu Leu Tyr Phe Lys Gln Leu Gly Ala Lys Leu Ile Glu Ala Ser Leu 50 55 60

Asp Asp His Gln Arg Leu Val Asp Ala Leu Lys Gin Val Asp Val Val 65 70 75 80

Ile Ser Ala Leu Ala Gly Val Leu Ser His His Ile Leu Glu Gln 85 90 95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe 100 105 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Asp Ile Met Glu His Ala Leu 115 120 125

-123-

-124-

Gln Pro Gly Ser Ile Thr Phe Ile Asp Lys Arg Iys Val Arg Arg Ala
 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310

Ile Glu Ala Ala Ser Ile Pro Tyr Thr Tyr Val Ser Ser Asn Met Phe
 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310

Ala Gly Tyr Phe Ala Gly Ser Leu Ala Glu Leu Asp Gly His Met Met
 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310

Pro Pro Arg Asp Lys Val Leu Ile Tyr Gly Asp Gly Asn Val Lys Gly
 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310

Ile Trp Val Asp Asp Val Gly Thr Tyr Ile Lys Ser Ile
 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310

Asp Asp Pro Gln Thr Leu Asn Lys Thr Met Tyr Ile Arg Pro Pro Met
 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310

Asn Ile Leu Ser Gln Lys Glu Val Ile Glu Ile Trp Glu Arg Ieu Ser
 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310

Glu Gln Asn Leu Asp Lys Ile Tyr Ile Ser Ser Gln Asp Phe Ieu Ala
 245 250 255 260 265 270 275 280 285 290 295 300 305 310

Asp Met Lys Asp Lys Ser Tyr Glu Gln Ile Val Arg Cys His Leu
 260 265 270 275 280 285 290 295 300 305 310

Tyr Gln Ile Phe Phe Arg Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro
 275 280 285 290 295 300 305 310

Asn Ala Ile Glu Ala Thr Ilys Leu Tyr Pro Glu Val Lys Tyr Val Thr
 290 295 300 305 310

Met Asp Ser Tyr Leu Glu Arg Tyr Val

ATG GAA GAG AGT AGC AGG GTT TTT ATA GTG GGA GGC ACA CGA TAC ATA
 Met Glu Glu Ser Ser Arg Val Lys Ile Val Gly Thr Gly Tyr Ile 108
 315 320 325

GCC AGA AGG ATT GTG AAA GCC AGC ATT GCT CTG GGC CAT CCT ACT TTC
 Gly Arg Arg Ile Val Lys Ala Ser Ile Ala Leu Gly His Pro Thr Phe 156
 330 335 340 345

ATT TTG TTT AGG AAA GAA GTT GTT TCT GAT GTA GAG AAA GTG GAG ATG
 Ile Leu Phe Arg Lys Glu Val Val Ser Asp Val Glu Lys Val Glu Met 204
 350 355 360

TTA TTG TCC TTC AAA AGA AAT GGT GGC AAA TTA CTG GNG GCT TCA TTT
 Leu Leu Ser Phe Lys Asn Gly Ala Lys Leu Glu Leu Ala Ser Phe 252
 365 370 375

GAT GAT CAC GAA AGC CTT GTA GAT CCT GTG AAG CAG GTT GAT GTT GTG
 Asp Asp Asp His Glu Ser Leu Val Asp Ala Val Lys Glu Val Asp Val Val 300
 380 385 390

ATA AGT GCA GTT GCA GGA AAC CAC ATG CGG CAT CAC ATC CTT CAA CGG
 Ile Ser Ala Val Ala Gly Asn His Met Arg His His Ile Leu Glu Glu 348
 395 400 405

CTC AAA TTA GTG GAG GCC ATT AAA GAA GCT GGA ATT ATT AAG AGG TTT
 Leu Lys Leu Val Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe 396
 410 415 420 425

GTT CCT TCA GAA TTT GGG ATG GAT CCA GGG TTA ATG GAG CAT GCA ATG
 Val Pro Ser Glu Phe Gly Met Asp Pro Gly Leu Met Glu His Ala Met 444
 430 435 440

GCA CCT GGC AAC ATT GTA TTT ATT GAT AAA ATA AAA GTT CGA GAG GCC
 Ala Pro Glu Asn Ile Val Phe Ile Asp Lys Ile Lys Val Arg Glu Ala 492
 445 450 455

ATA GAA GCT GCA TCC ATT CCT CAC ACT ATT ATC TCT GCC AAC ATA TTT
 Ile Glu Ala Ala Ser Ile Pro Ile Tyr Ile Ser Ala Asn Ile Phe 540
 460 465 470

GCT GGC TAC TTG GTT GCT GGA TTA GCT CAA CTT CGT GTG ATG CCT
 Ala Gly Tyr Leu Val Gly Leu Ala Glu Leu Gly Arg Val Met Pro 588
 475 480 485

TGG GTA GAA GAT GAT GTT GGA ATT GTC ATT GTC AAA GCA ATT GAT
 Trp Val Asp Glu Asp Asp Val Gly Ile Tyr Glu Asp Gly Asn Val Lys Ala Val 636
 490 495 500 505

GAC CCT CAC ACC CTA AAT AAG ACT ATG TAC AGG CCA CCT TTG AAT
 Asp Pro His Thr Leu Asn Lys Thr Met Tyr Ile Arg Pro Pro Leu Asn 732
 525 530 535

ATT CCT TCT CAG ARG GAA GTG GTT GAA AAA TGG GAA ARG TTA TCA GGA
 Ile Leu Ser Glu Lys Glu Val Val Lys Trp Glu Lys Leu Ser Gly 780
 540 545 550

60

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1151 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thujia plicata cDNA PLR-Tp2

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 61..996

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GATAAGCCAGC ATTCTTCAC CAAAGTCGTG CGCCATTANA GGAATAGTTT GAAAGCCAG

-125-

-126-

AAG AGC TTA AAT AAA ATA ATT TCT GCA GAT TTT CTT GCA GGC
 Lys Ser Leu Asn Lys Ile Asn Ile Ser Val Glu Asp Phe Leu Ala Gly 828
 555
 560
 ATG GAA GGT CAA TCA TAT GCA GAG CAG ATT GGA ATA TCA CAT TAC
 Met Glu Gly Gln Ser Tyr Gly Glu Gln Ile Gly Ile Ser His Phe Tyr 876
 570
 575
 CAA ATG TTC TAT AGG GGT GAT CTT TAT ATT GAA CCT AAT
 Gln Met Phe Tyr Arg Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asn 924
 590
 595
 600
 GGA GTC GAA GCT TCC CAA CCT TAT CCA GCA GAA TAA TAT ACA ACA GTG
 Gly Val Glu Ala Ser Gln Leu Tyr Pro Glu Val Lys Tyr Thr Thr Val 972
 605
 610
 GAT TCA TAC ATG GAA CGC TAC CTA TGCATCTCTT CTTCAAGAG ATATCTTAAAT
 Asp Ser Tyr Met Glu Arg Tyr Leu 1026
 620
 625
 TTAATTAGG CTTCTTAAATA GTTTTATAT TTGTCATTA TGCTAATAAA AAATGGAGAG
 TATCTGATA ATAATATGAA CCAATCATAT TAAATATAT TGGGATATAA AAAAAGAAA
 1086
 1146
 1151

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Met Glu Glu Ser Ser Arg Val Leu Ile Val Gly Thr Gly Tyr Ile
 1 5 10 15
 Gly Arg Arg Ile Val Lys Ala Ser Ile Ala Leu Gly His Pro Thr Phe
 20 25 30
 Ile Leu Phe Arg Lys Glu Val Val Ser Asp Val Glu Lys Val Glu Met
 35 40 45
 Leu Leu Ser Phe Lys Lys Asn Gly Ala Lys Leu Glu Ala Ser Phe
 50 55 60
 Asp Asp His Glu Ser Leu Val Asp Ala Val Lys Gin Val Asp Val Val
 65 70 75 80
 Ile Ser Ala Val Ala Gly Asn His Met Arg His His Ile Leu Glu Glu
 85 90 95
 Leu Lys Leu Val Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe
 100 105 110

Val Pro Ser Glu Phe Gly Met Asp Pro Gly Leu Met Glu His Ala Met
 115 120
 Ala Pro Gly Asn Ile Val Phe Ile Asp Lys Ile Lys Val Arg Glu Ala
 130 135
 Ile Glu Ala Ala Ser Ile Pro His Thr Tyr Ile Ser Ala Asn Ile Phe
 145 150
 Ala Gly Tyr Leu Val Gly Gly Leu Ala Glu Ileu Gly Arg Val Met Pro
 165 170
 Pro Ser Glu Lys Val Ile Leu Tyr Gly Asp Gly Asn Val Lys Ala Val
 180 185
 Trp Val Asp Glu Asp Asp Val Gly Ile Tyr Thr Ile Lys Ala Ile Asp
 195 200 205
 Asp Pro His Thr Leu Asn Lys Thr Met Tyr Ile Arg Pro Pro Leu Asn
 210 215 220
 Ile Leu Ser Gln Lys Glu Val Val Glu Lys Trp Glu Lys Leu Ser Gly
 225 230 235
 Lys Ser Leu Asn Lys Ile Asn Ile Ser Val Glu Asp Phe Leu Ala Gly
 245 250 255
 Met Glu Gly Gln Ser Tyr Gly Glu Ile Gly Ile Ser His Phe Tyr
 260 265 270
 Gln Met Phe Tyr Arg Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asn
 275 280 285
 Gly Val Glu Ala Ser Gln Leu Tyr Pro Glu Val Lys Tyr Thr Thr Val
 290 295 300
 Asp Ser Tyr Met Glu Arg Tyr Leu
 305 310

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1308 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thujia plicata cDNA PLR-tp3
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 164..1105
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

-127-

PCT/US97/20391 -

-128-

AAAAACTCTT AGACTTATT TCATTTTAC CCAGTCATA AGTGTGTGTT GGGTCTCTTC 60
 AAAAAGGCC CCCTCTCGTT AGAGCAAG AACAGCATGC TAGATGATAT GTAGAACCA 120
 AAATGCCRAA AATTGACTG TGAACTGGA TGCACTAGAATG GAT AAC AGG
 Met Asp Lys Lys 315
 AGC AGA CTT CTA ATA GTG GGG GGT ACT GGT TTT ATA GGC AAA AGA ATT 223
 Ser Arg Val Leu Ile Val Gly Ile Gly Phe Ile Gly Lys Arg Ile 320
 330
 GTG AAG GCC AGT TTG GCT CTT GCT CCT ACT TAT GTC TTG TTC AGC
 Val Lys Ala Ser Leu Ala Leu Gly His Pro Thr Val Leu Phe Arg 335
 340
 CCA GAA GCC CTC TCT TAC ATT GAC AAA GTG CAG ATG TTG ATA TCC TMC 271
 Pro Glu Ala Leu Ser Tyr Ile Asp Lys Val Gln Met Leu Ile Ser Phe 350
 360
 AAG CAG CTT GGG GCC AAA CTT CTT GAG GCT TCA TTG GAT GAC CAC CAA 319
 Lys Gin Leu Gly Ala Lys Leu Glu Ala Ser Leu Asp Asp His 365
 370
 GGG CTT GTG GAT GTT GTC AAA CAA GCA GAT GTT GTG ATC AGT GCT GTT 415
 Gly Leu Val Asp Val Val Lys Glu Val Asp Val Val Ile Ser Val 385
 390
 TCA GGA ATT AAA GAA GCT GGC ATT AGA ATT GCA CTT GAC CAG CTC AAG CTA GTG 463
 Ser Gly Ile Val Val Arg His His Ile Leu Asp Gln Leu Lys Leu Val 400
 405
 GAG GCA ATT AAA GAA GCT GGC ATT AGA ATT GCA CTT CCT CCA GAA 511
 Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe Leu Pro Ser Glu 415
 420
 TTT GGG ATG GAC CCA GAT GAT CCA TTG GAA CCT GGT AAC 559
 Phe Gly Met Asp Pro Asp Val Val Glu Asp Pro Leu Glu Pro Gly Asn 430
 435
 ATT ACA TTC ATT GAT AAA AGA AAA GTT AGA CGT GCC ATT GAA GCA GCA 607
 Ile Thr Phe Ile Asp Lys Arg Lys Val Arg Arg Ala Ile Glu Ala Ala 445
 450
 ACC ATT CCT TAC ACA TAT GTG TCT TCA ATT ATG TTT GCT GGG TTC TTT 655
 Thr Ile Pro Tyr Thr Tyr Val Ser Ser Asn Met Phe Ala Gly Phe Phe 465
 470
 GCT GGA AGC TTA GCA CAA CTG CAA GAT GCT CCC CGC ATG ATG GCT GCT 703
 Ala Gly Ser Leu Ala Glu Leu Glu Asp Ala Pro Arg Met Met Pro Ala 480
 485
 CGA GAT AAA GTT CTC ATA TAT GGA GAT GGA ATT GAT GCT GTT TAT 751
 Arg Asp Lys Val Leu Ile Tyr Gly Asp Gly Asn Val Lys Glu Val Tyr 495
 500
 GTA GAT GAA GAT GAT GCT GGA ATA TAC ATA GTC AAA TCA ATT GAT GAT 799
 Val Asp Glu Asp Asp Ala Gly Ile Tyr Ile Val Lys Ser Ile Asp Asp 510
 515
 520

CCT CGG AGA CTC AAC AAG ACT GTC TAT ATC AGG CCA CCA ATG AAT ATA 847
 Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro Met Asn Ile 525
 530
 CTT TCA CAG AAA GAA GTC GTT GAA ATA TGG GAG AGA CTA TCA GGT TTG 895
 Leu Ser Gln Lys Glu Val Val Ile Trp Glu Arg Leu Ser Gly Leu 545
 550
 AGC CTA GAA AAA ATC TAC GTC GTC GAC CAA CTT CTT AAT ATG AAA 943
 Ser Leu Glu Lys Ile Tyr Val Ser Gln Asp Glu Leu Ile Asn Met Lys 560
 565
 GAT AAA TCT TAT GTG GAG AAG ATG GCA CGA TGT CAT CTC TAT CAT TTT 991
 Asp Lys Ser Tyr Val Glu Lys Met Ala Arg Cys His Leu Tyr His Phe 575
 580
 TTT ATC ABA GGG GAT CTT TAC ATT TTT GAA ATT GCA CCC AAT GCT ACT 1039
 Phe Ile Lys Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asn Ala Thr 590
 595
 GAA GGC AGA AAA CTT TAT CCA GAA GTC AAA TAC AGC ACC ATG GAT TCA 1087
 Glu Gly Thr Lys Leu Tyr Pro Glu Val Lys Tyr Thr Thr Met Asp Ser 605
 610
 TAT ATG GAG CGT TAT CTA TAGCTTAATAG ATTTTCTTAATATAATAGCT 1135
 Tyr Met Glu Arg Tyr Ile Leu 625
 TGAATATTC TATCTCAAT AAGAGTGTAT TCTATAATAAT TACCAACAC TGCTCTTTT
 ATAGATTACT TTTTTAATAG GTGGCTTTA TAAACATGT ATAAAAAAA TTGCAAAACAA 1195
 TATTTTAAA TTAGCAATAA TAACCACCTT TAAATAAAAA AAAAAGAAA AAA 1255
 (2) INFORMATION FOR SEQ ID NO:66:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 314 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:
 Met Asp Lys Ser Arg Val Leu Ile Val Gly Thr Gly Phe Ile 1
 5
 Gly Lys Arg Ile Val Lys Ala Ser Leu Ala Leu Gly His Pro Thr Tyr 10
 15
 Val Leu Phe Arg Pro Glu Ala Leu Ser Tyr Ile Asp Lys Val Glu Met 20
 25
 30
 35
 40
 45
 Leu Ile Ser Phe Lys Glu Leu Gly Ala Lys Leu Glu Ala Ser Leu 50
 55
 60
 Asp Asp His Gln Gly Leu Val Asp Val Val Lys Glu Val Asp Val Val 65
 70
 80

Ile Ser Ala Val Ser Gly Gly Leu Val Arg His His Ile Leu Asp Gln
85 90 95
Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe
100 105 110
Leu Pro Ser Glu Phe Gly Met Asp Pro Asp Val Val Glu Asp Pro Leu
115 120 125
Glu Pro Gly Asn Ile Thr Phe Ile Asp Lys Val Arg Arg Ala
130 135 140
Ile Glu Ala Ala Thr Ile Pro Tyr Thr Tyr Val Ser Ser Asn Met Phe
145 150 155 160
Ala Gly Phe Phe Ala Gly Ser Leu Ala Glu Leu Gin Asp Ala Pro Arg
165 170 175
Met Met Pro Ala Arg Asp Lys Val Ile Tyr Gly Asp Gly Asn Val
180 185 190 195
Lys Gly Val Tyr Val Asp Glu Asp Asp Ala Gly Ile Tyr Ile Val Lys
195 200 205 210
Ser Ile Asp Asp Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro
215 220 225 230
Pro Met Asn Ile Leu Ser Gin Lys Glu Val Val Glu Ile Trp Glu Arg
235 240 245 250
Leu Ser Gly Leu Ser Leu Glu Lys Ile Tyr Val Ser Glu Asp Gln Leu
255 260 265 270
Leu Asn Met Lys Asp Lys Ser Tyr Val Glu Lys Met Ala Arg Cys His
275 280 285 290
Leu Tyr His Phe Ile Lys Gly Asp Leu Tyr Asn Phe Glu Ile Gly
295 300 305 310

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 11..946
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:
 GAAAGCAGAG ATG GAA GAG AGT AGC AGG ATT TAC GTA GTG GGA GGC ACA
 49
 Met Glu Glu Ser Ser Arg Ile Leu Val Val Gly Glu Thr
 315 320 325
 GGA TAC ATA GGC AGA AGG ATT GTG AAA GGC AGC ATT GCT CTG GGC CAT
 97
 Gly Tyr Ile Gly Arg Arg Ile Val Lys Ala Ser Ile Ala Leu Gly His
 330 335 340
 CCT ACT TTC ATT TTG TTG AGG AAA GAA GTT TCT GAT GTA GAG AAA
 145
 Pro Thr Phe Ile Leu Phe Arg Lys Glu Val Val Ser Asp Val Glu Lys
 345 350 355
 GTC GAG ATG TTA TTG TCC TTC AAA AAG ATT GGT GCC AAA TTA CTG GAG
 193
 Val Glu Met Leu Ser Phe Lys Asn Gly Ala Lys Leu Leu Glu
 360 365 370 375
 GCT TCA TTT GAT GAT CAC GAA AGC CTT GTC GAT GCT GAG CAG GTC
 241
 Ala Ser Phe Asp His Glu Ser Ile Val Asp Ala Val Lys Glu Val
 380 385 390
 GAT GTT GTC ATA AGT GCA GTC GCA GCA AAC CAC ATG CGG CAT CAC ATC
 289
 ASP Val Val Ile Ser Ala Val Ala Gly Asn His Met Arg His His Ile
 395 400 405
 CTT CAA CAG CTC ARA TTA GTC GAG GCC ATT AAA GAA GCT GGA AAT ATT
 337
 Leu Glu Gln Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile
 410 415 420
 AAG AGG TTT GTC CCT TCA GAA TTT GGG ATG GAT CCA GGG TTA ATG GAC
 385
 Lys Arg Phe Val Pro Ser Glu Phe Gly Met Asp Pro Gly Leu Met Asp
 425 430
 CAT GCA ATG GCA CCA GGA AAC ATT GTC TTT ATT GAT AAA ATA AAA GTC
 433
 His Ala Met Ala Pro Gly Asn Ile Val Phe Ile Pro His Thr Tyr Ile Ser Ala
 440 445 450 455
 CGA GAG GCC ATT GAA GCT GCA GCT ATT CCT CAC ACT TAT ATT TCT GCC
 481
 Arg Glu Ala Ile Glu Ala Ala Ile Pro His Thr Tyr Ile Ser Ala
 460 465 470
 AAT ATA TTT GCT GGC TAC TAC GTT GGT GGA TTA GCT GAA CTT GGT CGT
 529
 Asn Ile Phe Ala Gly Tyr Leu Val Gly Leu Ala Glu Leu Gly Arg
 475 480 485
 GTG ATG CCT CCT TCA GAC AAA GTC ATT GCA TAT GGA GAT GGA AAT GTC
 577
 Val Met Pro Pro Ser Asp Lys Val Phe Leu Tyr Gly Asp Gly Asn Val
 490 495 500
 AAA GCT GTT TGG ATA GAT GAA GAT GTT GGA ATA TAC ACA ATC ARA
 625
 Lys Ala Val Trp Ile Asp Glu Glu Asp Val Gly Ile Tyr Thr Ile Lys
 505 510 515

(2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1287 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Thujopsis dolabrata cDNA P1R-TP4
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

-131-

-132-

GCA ATT GAT GAC CCT CGC ACC CTA ATT AAG ACT GTC TAC ATC AGG CCA
Ala Ile Asp Asp Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro 673
520 525

CCG TTG ATT GTC CTT TCC CGG AAG GAA GTG CTT GAA AAA TGG GAA AAA
Pro Leu Asn Val Leu Ser Gin Lys Glu Val Val Lys Trp Glu Lys 721
540 545 550

TTA TCA AGA AAG AGC TTG GAT AAA ATA TAT ATG TCT GTT GAG GAT TTT
Leu Ser Arg Lys Ser Leu Asp Lys Ile Tyr Met Ser Val Glu Asp Phe 769
555 560

CTC GCA GGC ATG GAA GGT CAA TCA TAT GGA GAG ATT GGA ATA TCA
Leu Ala GLY Met Glu GLY GLN Ser Tyr GLY GLN Ile GLY Ile Ser 817
570 575 580

CAT TTC TAT CAG ATG TTC TAT BAG GGG GAT CTT TAT AAT TTT GAA ATT
His Phe Tyr Glu Met Phe Tyr Lys GLY Asp Leu Tyr Asn Phe Glu Ile 865
585 590 595

GGA CCT ATT GGA GTC TCC CAA CTT TAC CCA GGA GTC AAA TAC
Gly Pro Asn Gly Val Glu Ala Ser Glu Leu Tyr Pro Gly Val Lys Tyr 913
600 605 610 615

AGA AGA GTG GAC TCA TAC ATG GAG CGC TAC CTA TGAATATTG CTTCATGAG
TGAATGAGA GTATCTAGAT AATAATATC AATTGATAAT ATTCACTAT CAGTGAGAT 966
620 625

ATATTTAAAT TCAATTAAAT GCTTTCTAAAT AGTTTTATAA TTGTCGATAA ATGCTAAATA
TATGATGAGA 1026
GACTTTTCC CTTAACTGCA ATGCTGCAAA TATTTTATAC AACACAGCTTA ATGCTTTTA
AGGTGAGAA ACTAAATATG GTTTGTATT ACATGAGAAA ACCATTATTG GATATTGAG 1146
ATTGTTATTA TTTGAAATGT TATGATTGAT ATTAAATTG AATTTGATTA TGAACATGT
TTTAAAGAAA AAAAAAAA A 1287

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Met Glu Glu Ser Ser Arg Ile Leu Val Val Gly Gly Thr Gly Tyr Ile
1 5 10 15

Gly Arg Arg Ile Val Lys Ala Ser Ile Ala Leu Gly His Pro Thr Phe
20 25 30

Ile Leu Phe Arg Lys Glu Val Val Ser Asp Val Glu Lys Val Glu Met
35 40 45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

(A) LENGTH: 312 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

(A) LENGTH: 312 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

Leu Leu Ser Phe Lys Lys Asn Gly Ala Lys Leu Glu Ala Ser Phe
50 55 60

Asp Asp His Glu Ser Ile Val Asp Ala Val Lys Glu Val Asp Val Val
65 70 75 80

Ile Ser Ala Val Ala Gly Asn His Met Arg His His Ile Leu Glu Gln
85 90 95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe
100 105 110

Val Pro Ser Glu Phe Gly Met Asp Pro Gly Leu Met Asp His Ala Met
115 120 125

Ala Pro Gly Asn Ile Val Phe Ile Asp Lys Ile Lys Val Arg Glu Ala
130 135 140

Ile Glu Ala Ala Ile Pro His Thr Tyr Ile Ser Ala Asn Ile Phe
145 150 155 160

Ala Gly Tyr Leu Val GLY GLY Leu Ala Glu Leu Gly Arg Val Met Pro
165 170 175

Pro Ser Asp Lys Val Phe Leu Tyr Gly Asp Gly Asn Val Lys Ala Val
180 185 190

Trp Ile Asp Glu Asp Val Gly Ile Tyr Thr Ile Lys Ala Ile Asp
195 200 205

Asp Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro Pro Leu Asn
210 215 220

Val Leu Ser Gln Lys Glu Val Val Glu Lys Trp Glu Lys Leu Ser Arg
225 230 235 240

Lys Ser Leu Asp Lys Ile Tyr Met Ser Val Glu Asp Phe Leu Ala Gly
245 250 255

Met Glu Gly Gln Ser Tyr Gly Glu Lys Ile Gly Ile Ser His Phe Tyr
260 265 270

Gln Met Phe Tyr Lys Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asn
275 280 285

Gly Val Glu Ala Ser Gln Ile Tyr Pro Gly Val Lys Tyr Thr Thr Val
290 295 300

Asp Ser Tyr Met Glu Arg Tyr Leu
305 310

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1202 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: *Tsuga heterophylla* cDNA PLR-TH1

(i.ii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE: CDS

(A) NAME/KEY: CDS

(B) LOCATION: 2..922

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

C AGA GTT CTA ATA GTG GGT GCG ACA GGA TAC ATA GGT AGA AAA TTT
Arg Val Leu Ile Val Gly Gly Thr Gly Tyr Ile Gly Arg Lys Phe
315 320

GTA AAA GCT AGC TTA GTC CTA GGC CAC CCA ACA TTC GTT TTG TCC AGG
Val Lys Ala Ser Leu Ala Leu Gly His Pro Thr Phe Val Leu Ser Arg
330 335

CCA GAA GTA GGG TTT GAC ATT GAG AAG GTG CAC ATG TTG CTC TCC TTC
Pro Glu Val Gly Phe Asp Ile Glu Lys Val His Met Leu Ser Phe
345 350

AAA CAA GCG GGT GCC AGA CTT TTG GAG GGT TCA TTG GAT TCC GAA
Lys Gln Ala Gly Ala Arg Leu Leu Gly Ser Phe Glu Asp Phe Glu
360 365

AGC CTT GTG GCA GCC TTG AAG CAG GTC GAT GTC ATA AGT GCA GTG
Ser Leu Val Ala Ile Lys Glu Val Asp Val Val Ile Ser Ala Val
380 385

GCA GGA AAC CAT TTC AGA AAC CTT ATA CTT CAA CGG CTT AAA TTG GTC
Ala Gly Asn His Phe Arg Asn Leu Ile Leu Glu Gln Leu Val
395 400

GAA GCC ATA ATA GAA GCT GGC AAC ATT AAG AGA TTT CTT CCT GTC GAA
Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe Leu Pro Ser Glu
410 415

TTT GGA ATG GAA CCA GAC CTC ATG GAG CAC GCT TTG GAA CCT GGT AAC
Phe Gly Met Glu Pro Asp Leu Met Glu His Ala Leu Glu Pro Gly Asn
425 430

GCT GTC TTC ATT GAT AAG AGA AAC GCT CGG CGC GCC ATT GAA GCA GCA
Ala Val Phe Ile Asp Lys Arg Lys Val Arg Arg Ala Ile Glu Ala
440 445

GCG ATT CCT TAC ACG TAT GTC TCT ATA ATT GCA AAT ATA TTT GCT GGG TAT TTA
Gly Ile Pro Tyr Thr Tyr Val Ser Ser Asn Ile Phe Ala Gly Tyr Leu
450 460

GCA GGA GGG TTG GCA CAA ATT GGC CGG CTT ATG CCT CCT CGT GAT GAA
Ala Gly Leu Ala Glu Ile Gly Arg Leu Met Pro Arg Asp Glu
475 480

GTA GTT ATC TAT GGA GAT GGT AAC GCT TAA GCT GTC TGG CGT GAC GAA
Val Val Ile Tyr Gly Asp Gly Val Lys Ala Val Val Trp Val Asp Glu
490 500

GAT GAT GTC GGA ATA TAC ACA CGT AAA ACA ATC GAT GAT CCA CGC ACT
Asp Asp Val Gly Ile Tyr Thr Leu Lys Thr Ile Asp Asp Pro Arg Thr
505 510

CTG AAC AAC ACT GTC TAT ATC AGG CCA CTC AAA AAT ATT CTC TCT CAG
Leu Asn Lys Tyr Ile Arg Pro Leu Lys Asn Ile Leu Ser Glu
520 525 530

AAG GAG CTT GTG GCA AAG TGG GAA AAA CTC TCA GGA AAG TGT TTG AAG
Lys Glu Leu Val Ala Lys Trp Glu Lys Leu Ser Gly Lys Cys Leu Lys
540 545 550

AAA ACA TAC ATT TCT GCT GAG GAT TTT CTT GCA GGC ATC GAA GAT CAA
Lys Thr Tyr Ile Ser Ala Glu Asp Phe Leu Ala Gly Ile Glu Asp Glu
555 560 565

CCT TAC GAA CAT CGG GTC CGG ATA TCT CGC TTC TAT CAA ATG TTT TAC
Pro Tyr Glu His Gln Val Gly Ile Ser His Phe Tyr Glu Met Phe Tyr
570 575 580

AGT GGA GAT CRC TAT ATT GAG ATT GGG CCA GAC GGT AGA GAA GCA
Ser Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asp Gly Arg Glu Ala
585 590 595

ACA GTG CTA TAC CCT GAA GTT CAA TAC ACT ACC ATG GAT TCT TAT TTG
Thr Val Leu Tyr Pro Glu Val Gly Tyr Thr Met Asp Ser Tyr Leu
600 605 610

AAG CGC TAC TTA TAAGGAGGT GAAGGTAAAT GTTCCTAGAC ATGATACCA
Lys Arg Tyr Leu

CGAGAAATAC CAGAAATCTT CTTTCAGAGT CAAATAATGG ATAAATAATT CAACATTAGT
TCCATCAGAA ATACCGGAAAT TTTCATCTG AGTCACATAA ATGATAAAT ATTACATTAT
774

TTAAGTTTA TTATCGAAA TAGGGCTGAA CGAATTGAAAT ATATATCAT CTGATATGG
CGGGCAGGT GTAAATTTGC AAGCTGTACA GTRACTACGT CTGTCGGCA AAGGCTCTAA
TATCGATATA ACTGATGTA AAGGTACCA TTTCGTTATA ACTATACGT ATTATTT
TGACAAAAAA AAAAAAAA

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 307 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:
Arg Val Leu Ile Val Gly Asp Gly Thr Gly Tyr Ile Gly Arg Lys Phe Val
1 5 10 15

-135-

Lys Ala Ser Leu Ala Leu Gly His Pro Thr Phe Val Leu Ser Arg Pro
 20 25 30
 Glu Val Gly Phe Asp Ile Glu Lys Val His Met Leu Leu Ser Phe Lys
 35 40 45
 Gln Ala Gly Ala Arg Leu Leu Glu Gly Ser Phe Glu Asp Phe Gln Ser
 50 55 60
 Leu Val Ala Ala Leu Lys Gln Val Asp Val Ile Ser Ala Val Ala
 65 70 75 80
 GLY Asn His Phe Arg Asn Leu Ile Leu Gln Leu Lys Leu Val Glu
 85 90 95
 Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe Leu Pro Ser Glu Phe
 100 105 110
 Gly Met Glu Pro Asp Leu Met Glu His Ala Leu Glu Pro Gly Asn Ala
 115 120 125
 Val Phe Ile Asp Lys Arg Lys Val Arg Arg Ala Ile Glu Ala Ala Gly
 130 135 140
 Ile Pro Tyr Thr Tyr Val Ser Ser Asn Ile Phe Ala Gly Tyr Leu Ala
 145 150 155
 Gly Gly Leu Ala Glu Ile Gly Arg Leu Met Pro Pro Arg Asp Glu Val
 165 170 175
 Val Ile Tyr Gly Asp Gly Asn Val Lys Ala Val Trp Val Asp Glu Asp
 180 185 190
 Asp Val Gly Ile Tyr Thr Leu Lys Thr Ile Asp Asp Pro Arg Thr Leu
 195 200 205
 Asn Lys Thr Val Tyr Ile Arg Pro Leu Lys Asn Ile Leu Ser Gln Lys
 210 215 220
 Glu Leu Val Ala Ile Lys Trp Glu Lys Leu Ser Gly Lys Cys Leu Lys
 225 230 235 240
 Tyr Tyr Ile Ser Ala Glu Asp Phe Leu Ala Gly Ile Glu Asp Gln Pro
 245 250 255
 Tyr Glu His Gln Val Gly Ile Ser His Phe Tyr Gln Met Phe Tyr Ser
 260 265 270
 GLY Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asp Gly Arg Glu Ala Thr
 275 280 285
 Val Leu Tyr Pro Glu Val Gln Tyr Thr Thr Met Asp Ser Tyr Leu Lys
 290 295 300
 Arg Tyr Ile
 305
 310 315
 320 325
 330 335 340
 345 350
 355 360 365
 370 375
 380 385 390
 395 400
 405 410
 415 420 425
 430 435
 440 445
 450 455 460

(2) INFORMATION FOR SEQ ID NO:71:

-136-

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1328 base Pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (iii) MOLECULE TYPE: Tsuga heterophylla cDNA, PLR-Th2
 (i.ii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 20..946
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:
 GAATTGGGCA CGAGCTTAC ATG AGC AGA GTT CTA ATA GTG GGT GGC ACA GGA
 52
 met Ser Arg Val Ile Val Gly Glu Thr Gly
 TAC ATA GGT AGA AAA TTT GTC AAA GCT AGC TTA GCT CTA GGC CAC CCA
 Tyr Ile Gly Arg Lys Phe Val Lys Ala Ser Leu Ala Leu Gly His Pro
 100
 ACA TTC GTT TTG TCC AGG CCA GAA GTC GGG TTT GTC ATT GAG AAG GTG
 Thr Phe Val Ile Ser Arg Pro Glu Val Gly Phe Asp Ile Glu Lys Val
 148
 335 340 345 350
 CAC ATG TTG CTC TCC TCC AAA CAA GCG GGT GGC AGA CTT TTG GAG GGT
 His Met Leu Leu Ser Phe Lys Glu Ala Arg Leu Leu Glu Lys Val
 196
 TCA TTT GAG GAT TTC CTA AGC CTT TTG GCA GCC TTG AAG CAG GAT GAT
 Ser Phe Glu Asp Phe Glu Ser Leu Val Ala Ala Leu Lys Glu Val Asp
 244
 GTT GTG ATA AGT GCA GTG GCA GGA AAC CAT TTC AGA AAC CTT ATA CTT
 Val Val Ile Ser Ala Val Ala Gly Asn His Phe Arg Asn Leu Ile Leu
 292
 385 390 395
 CAA CAG CTC AAA TTG GTG GAA GCC ATA AAA GAG GCT CGC AAC ATT AGG
 Glu Gln Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Arg Asn Ile Lys
 340
 400 405 410
 AGA TTT CCT CCT GAA TTT GGA ATC GAC CCA GAC CTC ATG GAG CAC
 Arg Phe Leu Pro Ser Glu Phe Glu Met Asp Pro Asp Leu Met Glu His
 388
 415 420 425
 GCT TTG GAA CCT GGT AAC GCT GTC TTC ATT GAT AAG AGA AAG GTT CGG
 Ala Leu Glu Pro Gly Asn Ala Val Phe Ile Asp Lys Arg Lys Val Arg
 436
 435 440 445
 CCC GCC ATT GAA GCA GCA GGC ATT CCT TAC ACG TAT GTC TCT TCA AAT
 Arg Ala Ile Glu Ala Ala Gly Ile Ser Ser Asn
 484
 450 455 460

-139-

Ala Thr Met Leu Tyr Pro Glu Val Gln Tyr Thr Thr Met Asp Ser Tyr
 290 295 300

Leu Lys Arg Tyr Leu

(2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 355 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA probe used to isolate *Forsythia intermedia* dirigent protein cDNA clone
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:
 AAGGACCTGG 1GTTCTACTT CCACGACATA CTTTCGAAG GGGATAATAA CAACATGCC
 60
 ACTGCCACCA TAGTGGGTG CCCCCATGG GGCAACAGAA CTGGCATGG CGTGGCCATTG
 AATTTGGTG ACCTATGTT GTTGAGAT CCCATTACT TAGCACCAA TCTGATTCA
 120
 CCCCCAGTGG GTCGGGCACA AGGGATGTC TCTATGTC AAAAAGTAC ATACATGCT
 180
 TGGCTGGGT TCTGATTT GTTCAATICA ACTAACTATG TTGGACCTT GAACTTGTCT
 240
 GGGCTGATC CATGTTGAA CAGACTAGG GACGTATCAG TCACTGGG ACCA
 300
 355

(2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: "PCR primer R20"
- (iii) HYPOTHETICAL: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:
 CAGCTATGAC CATGTTAGC

(2) INFORMATION FOR SEQ ID NO:75:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. An isolated protein from a lignan biosynthetic pathway selected from the group consisting of dirigent protein and pinoresinol/lariciresinol reductases.
2. An isolated protein of Claim 1 having the biological activity of dirigent protein.
3. An isolated protein of Claim 2 having the biological activity of dirigent protein from *Forsythia*.
4. An isolated protein of Claim 3 having the biological activity of dirigent protein from *Forsythia intermedia*.
5. An isolated protein of Claim 2 having the biological activity of dirigent protein from *Trigia*.
6. An isolated protein of Claim 5 having the biological activity of dirigent protein from *Trigia heterophylla*.
7. An isolated protein of Claim 2 having the biological activity of dirigent protein from *Thuya*.
8. An isolated protein of Claim 7 having the biological activity of dirigent protein from *Thuya plicata*.
9. An isolated protein of Claim 1 having the biological activity of dirigent protein selected from the group consisting of SEQ ID Nos: 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and 35.
10. An isolated protein of Claim 1 having the biological activity of pinoresinol/lariciresinol reductase.
11. An isolated protein of Claim 10 having the biological activity of pinoresinol/lariciresinol reductase from *Forsythia*.
12. An isolated protein of Claim 11 having the biological activity of pinoresinol/lariciresinol reductase from *Forsythia intermedia*.

13. An isolated protein of Claim 10 having the biological activity of pinoresinol/lariciresinol reductase from *Trigia*.

14. An isolated protein of Claim 13 having the biological activity of pinoresinol/lariciresinol reductase from *Trigia heterophylla*.

15. An isolated protein of Claim 10 having the biological activity of pinoresinol/lariciresinol reductase from *Thuya*.

16. An isolated protein of Claim 15 having the biological activity of pinoresinol/lariciresinol reductase from *Thuya plicata*.

17. An isolated protein of Claim 1 having the biological activity of pinoresinol/lariciresinol reductase selected from the group consisting of SEQ ID Nos:48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70 and 72.

18. An isolated nucleotide sequence encoding a dirigent protein.

19. An isolated nucleotide sequence encoding a dirigent protein from *Forsythia* species.

20. A nucleotide sequence of Claim 19 encoding a dirigent protein from *Forsythia intermedia*.

21. An isolated nucleotide sequence encoding a protein having the biological activity of SEQ ID No:13 or SEQ ID No:15.

22. An isolated nucleotide sequence of Claim 19 which encodes the amino acid sequence of SEQ ID No:13 or SEQ ID No:15.

23. An isolated nucleotide sequence of Claim 19 having the sequence of SEQ ID No:12 or SEQ ID No:14.

24. An isolated nucleotide sequence encoding a dirigent protein from a *Tsuga* species.

25. A nucleotide sequence of Claim 24 encoding a dirigent protein from *Tsuga heterophylla*.

26. An isolated nucleotide sequence encoding a protein having the biological activity of SEQ ID No:17 or SEQ ID No:19.
27. An isolated nucleotide sequence of Claim 24 which encodes the amino acid sequence of SEQ ID No:17 or SEQ ID No:19.
28. An isolated nucleotide sequence of Claim 24 having the sequence of SEQ ID No:16 or SEQ ID No:18.
29. An isolated nucleotide sequence encoding a dirigent protein from a *Thuya* species.
30. A nucleotide sequence of Claim 29 encoding a dirigent protein from *Thuya plicata*.
31. An isolated nucleotide sequence encoding a protein having the biological activity of any one of SEQ ID Nos:21, 23, 25, 27, 29, 31, 33 or 35.
32. An isolated nucleotide sequence of Claim 29 which encodes the amino acid sequence of any one of SEQ ID Nos:21, 23, 25, 27, 29, 31, 33 or 35.
33. An isolated nucleotide sequence of Claim 29 having the sequence of any one of SEQ ID Nos:20, 22, 24, 26, 28, 30, 32 or 34.
34. An isolated nucleotide sequence encoding a pinoresinol/laricresinol reductase from a *Forsythia* species.
35. A nucleotide sequence of Claim 34 encoding a pinoresinol/laricresinol reductase from *Forsythia intermedia*.
36. An isolated nucleotide sequence encoding a protein having the biological activity of any one of SEQ ID Nos:48, 50, 52, 54, 56 or 58.
37. An isolated nucleotide sequence of Claim 34 which encodes the amino acid sequence of any one of SEQ ID Nos:48, 50, 52, 54, 56 or 58.
38. An isolated nucleotide sequence of Claim 34 having the sequence of any one of SEQ ID Nos:47, 49, 51, 53, 55 or 57.
39. An isolated nucleotide sequence encoding a pinoresinol/laricresinol reductase from a *Thuya* species.
40. A nucleotide sequence of Claim 39 encoding a pinoresinol/laricresinol reductase from *Thuya plicata*.
41. An isolated nucleotide sequence encoding a protein having the biological activity of any one of SEQ ID Nos:62, 64, 66 or 68.
42. An isolated nucleotide sequence of Claim 39 which encodes the amino acid sequence of any one of SEQ ID Nos:62, 64, 66 or 68.
43. An isolated nucleotide sequence of Claim 39 having the sequence of any one of SEQ ID Nos:61, 63, 65 or 67.
44. An isolated nucleotide sequence encoding a pinoresinol/laricresinol reductase from a *Truga* species.
45. A nucleotide sequence of Claim 44 encoding a pinoresinol/laricresinol reductase from *Truga heterophylla*.
46. An isolated nucleotide sequence encoding a protein having the biological activity of SEQ ID No:70 or SEQ ID No:72.
47. An isolated nucleotide sequence of Claim 44 which encodes the amino acid sequence of SEQ ID No:70 or SEQ ID No:72.
48. An isolated nucleotide sequence of Claim 44 having the sequence of SEQ ID No:69 or SEQ ID No:71.
49. A replicable expression vector comprising a nucleotide sequence encoding a protein having the biological activity of a dirigent protein selected from the group consisting of SEQ ID Nos:13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and 35.
50. A replicable expression vector comprising a nucleotide sequence encoding a protein having the biological activity of a pinoresinol/laricresinol reductase selected from the group consisting of SEQ ID Nos:48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70 and 72.

51. A host cell comprising a vector of Claim 49.

52. A host cell comprising a vector of Claim 50.

53. A method of enhancing the expression of pinoresinol/lariciresinol reductase in a suitable host cell comprising introducing into the host cell an expression vector that comprises a nucleotide sequence encoding a protein having the biological activity of a protein selected from the group consisting of SEQ ID Nos:48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70 and 72.

54. A method of modifying the expression of pinoresinol/lariciresinol reductase in a suitable host cell comprising introducing into the host cell an expression vector that comprises a nucleotide sequence that expresses an RNA that is complementary to all or part of a nucleic acid molecule selected from the group consisting of SEQ ID Nos:47, 49, 51, 53, 55, 57, 61, 63, 65, 67, 69 and 71.

55. A method of enhancing the expression of dirigent protein in a suitable host cell comprising introducing into the host cell an expression vector that comprises a nucleotide sequence that expresses an RNA that is complementary to all or part of a nucleic acid molecule selected from the group having the biological activity of a protein selected from the group consisting of SEQ ID Nos:13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and 35.

56. A method of modifying the expression of dirigent protein in a suitable host cell comprising introducing into the host cell an expression vector that comprises a nucleotide sequence that expresses an RNA that is complementary to all or part of a nucleic acid molecule selected from the group consisting of SEQ ID Nos:12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34.

57. A method of producing optically-pure lignans comprising introducing into a host cell an expression vector that comprises a nucleotide sequence encoding a dirigent protein capable of directing a bimolecular phenoxy coupling reaction to produce an optically-pure lignan, and purifying the optically-pure lignan from the host cell.

58. The method of Claim 57 wherein the nucleotide sequence is selected from the group consisting of SEQ ID Nos:12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34.

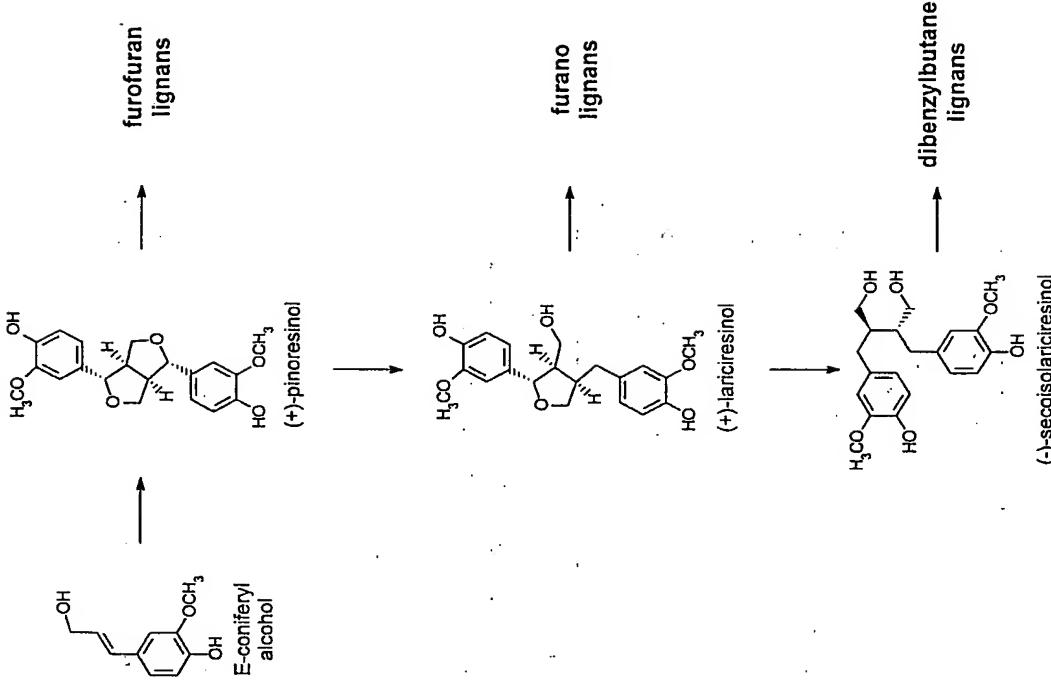


Fig. 1

INTERNATIONAL SEARCH REPORT

		International application No. PCT/US97/20391
<p>A. CLASSIFICATION OF SUBJECT MATTER</p> <p>IPC(6) :C12N 9/02, 15/53, 15/59 US CL :435/189, 252, 3, 325, 419, 320, 1; 536/23, 2, 23, 6; 530/370 According to International Patent Classification (IPC) or to both national classification and IPC</p>		
<p>B. FIELDS SEARCHED</p> <p>Minimum documentation searched (classification system followed by classification symbols)</p> <p>U.S. : 435/189, 252, 3, 325, 419, 320, 1; 536/23, 2, 23, 6; 530/370</p>		
<p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p> <p>NONE</p>		
<p>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)</p> <p>APS, DIALOG</p>		
<p>search terms: pinoresinol/lariciresinol reductase, dirigent protein, <i>Forsythia intermedia</i>, <i>Thujopsis dolabrata</i>, <i>Thuja heterophylla</i></p>		
<p>C. DOCUMENTS CONSIDERED TO BE RELEVANT</p>		
Category ^a	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHU et al. Stereospecificity of (+)-pinoresinol and (+)-lariciresinol reductases from <i>Forsythia intermedia</i> . The Journal of Biological Chemistry. 25 December 1993, Vol. 268, No. 36, pages 27026-27033, see entire document.	1,10-12 ----- 2-9, 13-58
A	KATAYAMA et al. An extraordinary accumulation of (-)-pinoresinol in cell-free extracts of <i>Forsythia intermedia</i> : Evidence for enantiospecific reduction of (+)-pinoresinol. Phytochemistry. 1992, Vol. 31, No. 11, pages 3875-3881, see entire document.	1-58
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
<p>^a Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document published on or after the international filing date</p> <p>"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of an invention or other special reason (e.g. specification)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>		
Date of the actual completion of the international search	Date of mailing of the international search report	
28 JANUARY 1998	23 FEB 1998	
Name and mailing address of the ISA/US Commissioner of Patent and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer KEITH D. HENDRICKS Telephone No. (703) 308-0196	

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.